CHEMOKINE-BINDING PROTEIN AND METHODS OF USE

RELATED APPLICATIONS

This application is a continuation-in-part of and claims priority under 35 U.S.C. § 120 to U.S. Patent Application Serial Number 10/317,832, entitled NOVEL DEATH ASSOCIATED PROTEINS, AND THAP1 AND PAR4 PATHWAYS IN APOPTOSIS CONTROL, filed December 10, 2002, which is a nonprovisional application of and claims benefit under 35 U.S.C. § 119(e) of U.S. Provisional Patent Application Serial Number 60/341,997, entitled NOVEL DEATH ASSOCIATED PROTEINS, AND THAP1 AND PAR4 PATHWAYS IN APOPTOSIS CONTROL, filed December 18, 2001. The disclosure of each of the above-listed priority applications is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

The present invention relates to genes and proteins of the THAP (THanatos (death)-Associated Protein) family, and uses thereof. In particular, the invention relates to uses of THAP-family proteins, or portions thereof, as chemokine-binding proteins and modulators of cellular and/or physiological responses.

BACKGROUND

Coordination of cell proliferation and cell death is required for normal development and tissue homeostasis in multicellular organisms. A defect in the normal coordination of these two processes is a fundamental requirement for tumorigenesis.

Progression through the cell cycle is highly regulated, requiring the transit of numerous checkpoints (for review, see Hunter, 1993). The extent of cell death is physiologically controlled by activation of a programmed suicide pathway that results in morphologically recognizable form of death termed apoptosis (Jacobson et al, 1997; Vaux et al., 1994). Both extra-cellular signals, such as tumor necrosis factor, and intracellular signals, like p53, can induce apoptotic cell death. Although many proteins involved in apoptosis or the cell cycle have been identified, the mechanisms by which these two processes are coordinated are not well understood.

It is well established that molecules which modulate apoptosis have the potential to treat a wide range of conditions relating to cell death and cell proliferation. For example, such molecules may be used for inducing cell death for the treatment of cancers, inhibiting cell death for the treatment of neurodegenerative disorders, and inhibiting or inducing cell death for regulating angiogenesis. However, because many biological pathways controlling cell cycle and apoptosis have not yet been fully elucidated, there is a need for the identification of biological targets for the development of therapeutic molecules for the treatment of these disorders.

PML nuclear bodies

PML nuclear bodies (PML-NBs), also known as PODs (PML oncogenic domains), ND10 (nuclear domain 10) and Kr bodies, are discrete subnuclear domains that are specifically disrupted in cells from acute promyelocytic leukemia (APL), a distinct subtype of human myeloid leukemia (Maul et al., 2000 ; Ruggero et al., 2000 ; Zhong et al., 2000a). Their name derives from their most intensively studied protein component, the promyelocytic leukemia protein (PML), a RING finger IFN-inducible protein encoded by a gene originally cloned as the t(15 ;17) chromosomal translocation partner of the retinoic acid receptor (RAR) locus in APL. In APL cells, the presence of the leukemogenic fusion protein, PML-RAR, leads to the disruption of PML-NBs and the delocalization of PML and other PML-NB proteins into aberrant nuclear structures (Zhong et al., 2000a). Treatment of both APL cell lines and patients with retinoic acid, which induces the degradation of the PML-RAR oncoprotein, results in relocalization of PML and other NBs components into PML-NBs and complete remission of clinical disease, respectively. The deregulation of the PML-NBs by PML-RAR thus appears to play a critical role in tumorigenesis. The analysis of mice, where the PML gene was disrupted by homologous recombination, has revealed that PML functions as a tumor suppressor *in vivo* (Wang et al., 1998a), that is essential for multiple apoptotic pathways (Wang et al., 1998b). Pml -/- mice and cells are protected from Fas, TNFα, ceramide and IFN-induced apoptosis as well as from DNA damage-induced apoptosis. However, the molecular mechanisms through which PML modulates the response to pro-apoptotic stimuli are not well understood (Wang et al., 1998b; Quignon et al., 1998). Recent studies indicate that PML can participate in both p53-dependent and p53-independent

apoptosis pathways (Guo et al., 2000; Fogal et al., 2000). p53-dependent DNA-damage induced apoptosis, transcriptional activation by p53 and induction of p53 target genes are all impaired in PML -/- primary cells (Guo et al., 2000). PML physically interacts with p53 and acts as a transcriptional co-activator for p53. This co-activatory role of PML is absolutely dependent on its ability to recruit p53 in the PML-NBs (Guo et al., 2000; Fogal et al., 2000). The existence of a cross-talk between PML- and p53-dependent growth suppression pathways implies an important role for PML-NBs and PML-NBs-associated proteins as modulators of p53 functions. In addition to p53, the pro-apoptotic factor Daxx could be another important mediator of PML pro-apoptotic activities (Ishov et al., 1999; Zhong et al., 2000b; Li et al., 2000). Daxx was initially identified by its ability to enhance Fas-induced cell death. Daxx interacts with PML and localizes preferentially in the nucleus where it accumulates in the PML-NBs (Ishov et al., 1999; Zhong et al., 2000b; Li et al., 2000). Inactivation of PML results in delocalization of Daxx from PML-NBs and complete abrogation of Daxx pro-apoptotic activity (Zhong et al., 2000b). Daxx has recently been found to possess strong transcriptional repressor activity (Li et al., 2000). By recruiting Daxx to the PML-NBs, PML may inhibit Daxx-mediated transcriptional repression, thus allowing the expression of certain pro-apoptotic genes.

PML-NBs contain several other proteins in addition to Daxx and p53. These include the autoantigens Sp100 (Sternsdorf et al., 1999) and Sp100-related protein Sp140 (Bloch et al., 1999), the retinoblastoma tumor suppressor pRB (Alcalay et al., 1998), the transcriptional co-activator CBP (LaMorte et al., 1998), the Bloom syndrome DNA helicase BLM (Zhong et al., 1999) and the small ubiquitin-like modifier SUMO-1 (also known as sentrin-1 or PIC1; for recent reviews see Yeh et al., 2000; Melchior, 2000; Jentsch and Pyrowolakis, 2000). Covalent modification of PML by SUMO-1 (sumoylation) appears to play a critical role in PML accumulation into NBs (Muller et al., 1998) and the recruitment of other NBs components to PML-NBs (Ishov et al., 1999; Zhong et al., 2000c).

Prostate apoptosis response-4

Prostate apoptosis response-4 (PAR4) is a 38 kDa protein initially identified as the product of a gene specifically upregulated in prostate tumor cells undergoing

apoptosis (for reviews see Rangnekar, 1998 ; Mattson et al., 1999). Consistent with an important role of PAR4 in apoptosis, induction of PAR4 in cultured cells is found exclusively during apoptosis and ectopic expression of PAR4 in NIH-3T3 cells (Diaz-Meco et al., 1996), neurons (Guo et al., 1998), prostate cancer and melanoma cells (Sells et al., 1997) has been shown to sensitize these cells to apoptotic stimuli. In addition, down regulation of PAR4 is critical for ras-induced survival and tumor progression (Barradas et al., 1999) and suppression of PAR4 production by antisense technology prevents apoptosis in several systems (Sells et al., 1997; Guo et al., 1998), including different models of neurodegenerative disorders (Mattson et al., 1999), further emphasizing the critical role of PAR4 in apoptosis. At the carboxy terminus, PAR4 contains both a leucine zipper domain (Par4LZ, amino acids 290-332), and a partially overlapping death domain (Par4DD, amino acids 258-332). Deletion of this carboxy-terminal part abrogates the pro-apoptotic function of PAR4 (Diaz-Meco et al., 1996 ; Sells et al., 1997 ; Guo et al., 1998). On the other hand, overexpression of PAR4 leucine zipper/death domain acts in a dominant negative manner to prevent apoptosis induced by full-length PAR4 (Sells et al., 1997 ; Guo et al., 1998). The PAR4 leucine zipper/death domain mediates PAR4 interaction with other proteins by recognizing two different kinds of motifs : zinc fingers of the Wilms tumor suppressor protein WT1 (Johnstone et al., 1996) and the atypical isoforms of protein kinase C (Diaz-Meco et al., 1996), and an arginine-rich domain from the death-associated-protein (DAP)-like kinase Dlk (Page et al., 1999). Among these interactions, the binding of PAR4 to aPKCs and the resulting inhibition of their enzymatic activity is of particular functional relevance because the aPKCs are known to play a key role in cell survival and their overexpression has been shown to abrogate the ability of PAR4 to induce apoptosis (Diaz-Meco et al., 1996 ; Berra et al., 1997).

SLC/CCL21

Chemokine SLC/CCL21 (also known as SLC, CKβ-9, 6Ckine, and exodus-2) is a member of the CC (beta)-chemokine subfamily. SLC/CCL21 contains the four conserved cysteines characteristic of beta chemokines plus two additional cysteines in its unusually long carboxyl-terminal domain. Human SLC/CCL21 cDNA encodes a 134 amino acid residue, highly basic, precursor protein with a 23 amino acid residue signal

peptide that is cleaved to form the predicted 111 amino acid residues mature protein. Mouse SLC/CCL21 cDNA encodes a 133 amino acid residue protein with 23 residue signal peptide that is cleaved to generate the 110 residue mature protein. Human and mouse SLC/CCL21 is highly conserved, exhibiting 86% amino acid sequence identity. The gene for human SLC/CCL21 has been localized at human chromosome 9p13 rather than chromosome 17, where the genes of many human CC chemokines are clustered. The SLC/CCL21 gene location is within a region of about 100 kb as the gene for MIP-3 beta/ELC/CCL19, another recently identified CC chemokine. SLC/CCL21 was previously known to be highly expressed in lymphoid tissues at the mRNA level, and to be a chemoattractant for T and B lymphocytes (Nagira, et al. (1997) J. Biol. Chem. 272:19518-19524; Hromas, et al. (1997) J. Immunol. 159:2554-2558; Hedrick, et al. (1997) J. Immunol. 159:1589-1593; Gunn, et al. (1998) Proc. Natl. Acad. Sci. 95:258-263). SLC/CCL21 also induces both adhesion of lymphocytes to intercellular adhesion molecule-1 and arrest of rolling cells (Campbell, et al. (1998) Science 279:381-384). All of the above properties are consistent with a role for SLC/CCL21 in regulating trafficking of lymphocytes through lymphoid tissues. Unlike most CC chemokines, SLC/CCL21 is not chemotactic for monocytes. However, it has been reported to inhibit hemopoietic progenitor colony formation in a dose-dependent manner (Hromas et al. (1997) J. Immunol. 159: 2554-58).

Chemokine SLC/CCL21 is a ligand for chemokine receptor CCR7 (Rossi et al. (1997) J. Immunol. 158:1033; Yoshida et al. (1997) J. Biol. Chem. 272:13803; Yoshida et al. (1998) J. Biol. Chem. 273:7118; Campbell et al. (1998) J Cell Biol 141:1053). CCR7 is expressed on T cells and dendritic cells (DC), consistent with the chemotactic action of SLC/CCL21 for both lymphocytes and mature DC. Both memory ($CD45RO^+$) and naïve ($CD45RA^+$) $CD4^+$ and $CD8^+$ T cells express the CCR7 receptor (Sallusto et al. (1999) Nature 401:708). Within the memory T cell population, CCR7 expression discriminates between T cells with effector function that can migrate to inflamed tissues ($CCR7^-$) vs. T cells that require a secondary stimulus prior to displaying effector functions ($CCR7^+$) (Sallusto et al. (1999) Nature 401:708). Unlike mature DC, immature DC do not express CCR7 nor do they respond to the chemotactic action of CCL21 (Sallusto et al. (1998) Eur. J. Immunol. 28:2760; Dieu et al. (1998) J. Exp. Med. 188:373).

A key function of CCR7 and its two ligands SLC/CCL21 and MIP3b/CCL19 is facilitating recruitment and retention of cells to secondary lymphoid organs in order to promote efficient antigen exposure to T cells. CCR7-deficient mice demonstrate poorly developed secondary organs and exhibit an irregular distribution of lymphocytes within lymph nodes, Peyer's patches, and splenic periarteriolar lymphoid sheaths (Forster et al. (1999) Cell 99:23). These animals have severely impaired primary T cell responses largely due to the inability of interdigitating DC to migrate to the lymph nodes (Forster et al. (1999) Cell 99:23). The overall findings to date support the notion that CCR7 and its two ligands, CCL19 and CCL21, are key regulators of T cell responses via their control of T cell/DC interactions. CCR7 is an important regulatory molecule with an instructive role in determining the migration of cells to secondary lymphoid organs (Forster et al. (1999) Cell 99:23; Nakano et al. (1998) Blood 91:2886).

SUMMARY OF THE INVENTION

THAP1 (THanatos-Associated-Protein-1)

In the past few years, the inventors have focused on the molecular characterization of novel genes expressed in the specialized endothelial cells (HEVECs) of post-capillary high endothelial venules (Girard and Springer, 1995a; Girard and Springer, 1995b; Girard et al., 1999). In the present invention, they report the analysis of THAP1 (for THanatos (death)-Associated Protein-1), a protein that localizes to PML-NBs. Two hybrid screening of an HEVEC cDNA library with the THAP1 bait lead to the identification of a unique interacting partner, the pro-apoptotic protein PAR4. PAR4 is also found to accumulate into PML-NBs and targeting of the THAP1 / PAR4 complex to PML-NLs is mediated by PML. Similarly to PAR4, THAP1 is a pro-apoptotic polypeptide. Its pro-apoptotic activity requires a novel protein motif in the amino-terminal part called THAP domain. Together these results define a novel PML-NBs pathway for apoptosis that involves the THAP1/PAR4 pro-apoptotic complex.

Embodiments of the present invention include genes, proteins and biological pathways involved in apoptosis. In some embodiments, the genes, proteins, and pathways disclosed herein may be used for the development of polypeptide, nucleic acid or small molecule therapeutics.

One embodiment of the present invention provides a novel protein motif, the THAP domain. The present inventors initially identified the THAP domain as a 90 residue protein motif in the amino-terminal part of THAP1 and which is essential for THAP1 pro-apoptotic activity. THAP1 (THanatos (death) Associated Protein-1), as determined by the present inventors, is a pro-apoptotic polypeptide which forms a complex with the pro-apoptotic protein PAR4 and localizes in discrete subnuclear domains known as PML nuclear bodies. However, the THAP domain also defines a novel family of proteins, the THAP family, and the inventors have also provided at least twelve distinct members in the human genome (THAP-0 to THAP11), all of which contain a THAP domain (typically 80-90 amino acids) in their amino-terminal part. The present invention thus includes nucleic acid molecules, including in particular the complete cDNA sequences, encoding members of the THAP family, portions thereof encoding the THAP domain or polypeptides homologous thereto, as well as to polypeptides encoded by the THAP family genes. The invention thus also includes diagnostic and activity assays, and uses in therapeutics, for THAP family proteins or portions thereof, as well as drug screening assays for identifying compounds capable of inhibiting (or stimulating) pro-apoptotic activity of a THAP family member.

In one example of a THAP family member, THAP1 is determined to be an apoptosis inducing polypeptide expressed in human endothelial cells (HEVECs), providing characterization of the THAP sequences required for apoptosis activity in the THAP1 polypeptide. In further aspects, the invention is also directed to the interaction of THAP1 with the pro-apoptotic protein PAR4 and with PML-NBs, including methods of modulating THAP1 / PAR4 interactions for the treatment of disease. The invention also concerns interaction between PAR4 and PML-NBs, diagnostics for detection of said interaction (or localization) and modulation of said interactions for the treatment of disease.

Compounds which modulate interactions between a THAP family member and a THAP-family target molecule, a THAP domain or THAP-domain target molecule, or a PAR4 and a PML-NBs protein may be used in inhibiting (or stimulating) apoptosis of different cell types in various human diseases. For example, such compounds may be used to inhibit or stimulate apoptosis of endothelial cells in angiogenesis-dependent diseases including but not limited to cancer, cardiovascular diseases, inflammatory

diseases, and to inhibit apoptosis of neurons in acute and chronic neurodegenerative disorders, including but not limited to Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral sclerosis, HIV encephalitis, stroke, epileptic seizures).

Oligonucleotide probes or primers hybridizing specifically with a THAP1 genomic DNA or cDNA sequence are also part of the present invention, as well as DNA amplification and detection methods using said primers and probes.

Fragments of THAP family members or THAP domains include fragments encoded by nucleic acids comprising at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 consecutive nucleotides selected from the group consisting of SEQ ID NOs: 160-175, or polypeptides comprising at least 8, 10, 12, 15, 20, 25, 30, 40, 50, 100, 150 or 200 consecutive amino acids selected from the group consisting of SEQ ID NOs: 1-114.

A further aspect of the invention includes recombinant vectors comprising any of the nucleic acid sequences described above, and in particular to recombinant vectors comprising a THAP1 regulatory sequence or a sequence encoding a THAP1 protein, THAP family member, THAP domain, fragments of THAP family members and THAP domains, homologues of THAP family members/ THAP domains, as well as to cell hosts and transgenic non human animals comprising said nucleic acid sequences or recombinant vectors.

Another aspect of the invention relates to methods for the screening of substances or molecules that inhibit or increase the expression of the THAP1 gene or genes encoding THAP family members, as well as with methods for the screening of substances or molecules that interact with and/or inhibit or increase the activity of a THAP1 polypeptide or THAP family polypeptide.

In accordance with another aspect, the present invention provides a medicament comprising an effective amount of a THAP family protein, e. g. THAP1, or a SLC/CCL21-binding fragment thereof, together with a pharmaceutically acceptable carrier. The medicaments described herein may be useful for treatment and/or prophylaxis.

As related to another aspect, the invention is concerned in particular with the use of a THAP family protein, homologs thereof and fragments thereof, for example THAP1, or a SLC/CCL21-binding fragment thereof as an anti-inflammatory agent. The

THAP family protein, for example, THAP1 and fragments thereof will be useful for the treatment of conditions mediated by SLC/CCL21.

In a further aspect, the present invention provides a detection method comprising the steps of providing a SLC/CCL21 chemokine-binding molecule which is a THAP family protein, for example, THAP1, or an SLC/CCL21-binding fragment thereof, contacting the SLC/CCL21-binding THAP1 molecule with a sample, and detecting an interaction of the SLC/CCL21-binding THAP1 molecule with SLC/CCL21 chemokine in the sample.

In one example, the invention may be used to detect the presence of SLC/CCL21 chemokine in a biological sample. The SLC/CCL21-binding THAP1 molecule may be usefully immobilized on a solid support, for example as a THAP1/Fc fusion.

In accordance with another aspect, the present invention provides a method for inhibiting the activity of SLC/CCL21 chemokine in a sample, which method comprises contacting the sample with an effective amount of a SLC/CCL21 chemokine-binding molecule which is a THAP1 protein or a SLC/CCL21-binding fragment thereof.

In further aspects the invention provides a purified THAP1 protein or a SLC/CCL21-binding fragment thereof, or a THAP1/Fc fusion, for use in a method or a medicament as described herein; and a kit comprising such a purified THAP1 protein or fragment.

Some embodiments of the invention also envisage the use of fragments of the THAP1 protein, which fragments have SLC/CCL21 chemokine-binding properties. The fragments may be peptides derived from the protein. Use of such peptides can be preferable to the use of an entire protein or a substantial part of a protein, for example because of the reduced immunogenicity of a peptide compared to a protein. Such peptides may be prepared by a variety of techniques including recombinant DNA techniques and synthetic chemical methods.

In addition to the above properties, THAP1 has the capability to bind to several additional chemokines. Such chemokines include, but are not limited to, ELC/CCL19, RANTES CCL5, MIG/CXCL9 and IP10/CXCL10. As such, further aspects of the present invention relate to the binding of chemokines by THAP1, a chemokine binding domain of THAP1, and polypeptides having at least 30% amino acid identity to THAP1 or a chemokine-binding domain of THAP1. Also contemplated is the binding of

chemokines to oligomers and Fc immunoglobulin fusions of the above-listed polypeptides.

According to some aspects of the present invention, a THAP1 polypeptide, a chemokine-binding domain of THAP1, polypeptides having at least 30% amino acid identity to THAP1 or a chemokine-binding domain of THAP1 as well as oligomers or Fc immunoglobulin fusions of these proteins can be used in pharmaceutical compositions and/or medicaments for reducing the symptoms associated with inflammation and/or inflammatory diseases. As such, some aspects of the present invention include pharmaceutical compositions and/or medicaments comprising THAP1 protein, a chemokine-binding domain of THAP1, polypeptides having at least 30% amino acid identity to THAP1 or a chemokine-binding domain of THAP1 as well as oligomers or Fc immunoglobulin fusions of these proteins.

Yet other aspects of the invention relate THAP-family polypeptides, chemokine binding domains of THAP-family peptides, fusions of a THAP-family polypeptide with an immunoglobulin Fc region, fusions of a chemokine-binding domain of a THAP-family peptide with an immunoglobulin Fc region, oligomers of THAP family polypeptides, chemokine-binding domains of THAP family peptides, THAP-family peptide-Fc fusions, and chemokine-binding domain of THAP-family peptide-Fc fusions as well as polypeptides having at least 30% amino acid identity to any of the above-listed polypeptides. Pharmaceutical compositions which include one or more of these polypeptides are also contemplated.

Aspects of the invention relate to methods of binding a chemokine, inhibiting the activity of a chemokine, reducing or ameliorating the symptoms of a condition mediated or influenced by one or more chemokines, preventing the symptoms of a condition mediated or influenced by one or more chemokines and detecting a chemokine by using a chemokine-binding agents such as THAP-family polypeptides, chemokine binding domains of THAP-family peptides, fusions of a THAP-family polypeptide with an immunoglobulin Fc region, fusions of a chemokine-binding domain of a THAP-family peptide with an immunoglobulin Fc region, oligomers of THAP family polypeptides, chemokine-binding domains of THAP family peptides, THAP-family peptide-Fc fusions, and chemokine-binding domain of THAP-family peptide-Fc fusions as well as

polypeptides having at least 30% amino acid identity to any of the above-listed polypeptides.

It will also be evident that the THAP-family proteins for use in the invention may be prepared in a variety of ways, in particular as recombinant proteins in a variety of expression systems. Any standard systems may be used such as baculovirus expression systems or mammalian cell line expression systems.

Other aspects of the invention are described in the following numbered paragraphs:

1. A method of identifying a candidate modulator of apoptosis comprising:

(a) contacting a THAP-family polypeptide or a biologically active fragment thereof with a test compound, wherein said THAP-family polypeptide comprises at least 30% amino acid identity to an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114; and

(b) determining whether said compound selectively modulates the activity of said polypeptide;

wherein a determination that said test compound selectively modulates the activity of said polypeptide indicates that said compound is a candidate modulator of apoptosis.

2. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 3, or a biologically active fragment thereof.

3. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 4, or a biologically active fragment thereof.

4. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 5, or a biologically active fragment thereof.

5. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 6, or a biologically active fragment thereof.

6. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 7, or a biologically active fragment thereof.

7. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 8, or a biologically active fragment thereof.

8. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 9, or a biologically active fragment thereof.

9. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 10, or a biologically active fragment thereof.

10. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 11, or a biologically active fragment thereof.

11. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 12, or a biologically active fragment thereof.

12. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 13, or a biologically active fragment thereof.

13. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 14, or a biologically active fragment thereof.

14. The method of Paragraph 1, wherein the THAP-family polypeptide comprises the amino acid sequence selected from the group consisting of SEQ ID NOs: 15-114, and biologically active fragments thereof.

15. The method of Paragraph 1, wherein said biologically active fragment of said THAP-family protein has at least one biological activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis.

16. The methods of any one of Paragraphs 2-15 wherein said THAP-family polypeptide has at least one biological activity selected from the group consisting of

interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis.

17. An isolated nucleic acid encoding a polypeptide having apoptotic activity, said polypeptide consisting essentially of an amino acid sequence selected from the group consisting of :

(a) amino acid positions 1-90 of SEQ ID NO: 2, a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto ;

(b) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 89 of SEQ ID NO: 3, a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto ;

(c) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 89 of SEQ ID NO: 4, a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto ;

(d) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 89 of SEQ ID NO: 5, a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto ;

(e) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 90 of SEQ ID NO: 6, a fragment thereof having apoptotic activity or a polypeptide having at least 30% amino acid identity thereto ;

(f) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 90 of SEQ ID NO: 7, a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto ;

(g) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 90 of SEQ ID NO: 8, a fragment thereof

having apoptotic activity ; or a polypeptide having at least 30% amino acid identity thereto ;

(h) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 90 of SEQ ID NO: 9, a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto ;

(i) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 92 of SEQ ID NO: 10, a fragment thereof having apoptotic activity or a polypeptide having at least 30% amino acid identity thereto ;

(j) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 90 of SEQ ID NO: 11, a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto ;

(k) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 90 of SEQ ID NO: 12, or a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto ;

(l) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 90 of SEQ ID NO: 13, a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto ; and

(m) a polypeptide comprising a THAP-family domain consisting essentially of amino acid positions 1 to 90 of SEQ ID NO: 14, a fragment thereof having apoptotic activity, or a polypeptide having at least 30% amino acid identity thereto.

18. An isolated nucleic acid encoding a THAP-family polypeptide having apoptotic activity selected from the group consisting of:

(i) a nucleic acid molecule encoding a polypeptide comprising the amino acid sequence of a sequence selected from the group consisting of SEQ ID NOs: 1-114;

(ii) a nucleic acid molecule comprising the nucleic acid sequence of a sequence selected from the group consisting of SEQ ID NOs: 160-175 and the sequences complementary thereto; and

(iii) a nucleic acid the sequence of which is degenerate as a result of the genetic code to the sequence of a nucleic acid as defined in (i) and (ii).

19. The nucleic acid of Paragraph 18, wherein said nucleic acid comprises a nucleic acid selected from the group consisting of SEQ ID NOs. 5, 7, 8 and 11.

20. The nucleic acid of Paragraph 18, wherein said nucleic acid comprises a nucleic acid selected from the group consisting of SEQ ID NOs. 162, 164, 165 and 168.

21. An isolated nucleic acid encoding a THAP-family polypeptide having apoptotic activity comprising:

(i) the nucleic acid sequence of SEQ ID NOs : 1-2 or the sequence complementary thereto ; or

(ii) a nucleic acid molecule encoding a polypeptide comprising the amino acid sequence of SEQ ID NOs 1-2;

22. An isolated nucleic acid, said nucleic acid comprising a nucleotide sequence encoding:

i) a polypeptide comprising an amino acid sequence having at least about 80% identity to a sequence selected from the group consisting of the polypeptides of SEQ ID NOs: 1-114 and the polypeptides encoded by the nucleic acids of SEQ ID NOs: 160-175 or

ii) a fragment of said polypeptide which possesses apoptotic activity.

23. The nucleic acid of Paragraph of Paragraph 23, wherein said nucleic acid encodes a polypeptide comprising an amino acid sequence having at least about 80% identity to a sequence selected from the group consisting of the polypeptides of SEQ ID NOs: 5, 7, 8 and 11 and the polypeptides encoded by the nucleic acids of SEQ ID NOs: 162, 164, 165 and 168 or a fragment of said polypeptide which possesses apoptotic activity.

24. The nucleic acid of Paragraph 23, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of the sequences of SEQ ID NOs: 5, 7, 8 and 11 and the polypeptides encoded by the nucleic acids of SEQ ID NOs: 162, 164, 165 and 168.

25. The nucleic acid of Paragraph 23, wherein polypeptide identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the default parameters of XBLAST.

26. The nucleic acid of Paragraph 17, wherein said nucleic acid is operably linked to a promoter.

27. An expression cassette comprising the nucleic acid of Paragraph 26.

28. A host cell comprising the expression cassette of Paragraph 27.

29. A method of making a THAP-family polypeptide, said method comprising

providing a population of host cells comprising a recombinant nucleic acid encoding said THAP-family protein of any one of SEQ ID NOs. 1-114; and

culturing said population of host cells under conditions conducive to the expression of said recombinant nucleic acid;

whereby said polypeptide is produced within said population of host cells.

30. The method of Paragraph 29 wherein said providing step comprises providing a population of host cells comprising a recombinant nucleic acid encoding said THAP-family protein of any one of SEQ ID NOs. 5, 7, 8 and 11.

31. The method of Paragraph 29, further comprising purifying said polypeptide from said population of cells.

32. An isolated THAP polypeptide encoded by the nucleic acid of any one of SEQ ID Nos. 160-175.

33. The polypeptide of Paragraph 32, wherein said polypeptide is encoded by a nucleic acid selected from the group consisting of SEQ ID NOs. 5, 7, 8, 11, 162, 164, 165 and 168.

34. The polypeptide of Paragraph 32, wherein said polypeptide has at least one activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis.

35. An isolated THAP polypeptide or fragment thereof, said polypeptide comprising at least 12 contiguous amino acids of a sequence selected from the group consisting of SEQ ID NOs: 1-114.

36. The polypeptide of Paragraph 35, wherein said polypeptide comprises at least 12 contiguous amino acids of a sequence selected from the group consisting of SEQ ID NOs. 5, 7, 8, and 11.

37. The polypeptide of Paragraph 35, wherein said polypeptide has at least one activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis.

38. An isolated THAP polypeptide or fragment thereof, said polypeptide comprising an amino acid sequence having at least about 80% amino acid sequence identity to a sequence selected from the group consisting of SEQ ID NOs: 1-114 or a fragment thereof, said polypeptide or fragment thereof having at least one activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis.

39. The polypeptide of Paragraph 38, wherein said THAP polypeptide or fragment thereof comprises an amino acid sequence having at least about 80% amino acid sequence identity to a sequence selected from the group consisting of SEQ ID NOs: 5, 7, 8 and 11 or a fragment thereof having at least one activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis.

40. The polypeptide of Paragraph 38, wherein said polypeptide is selectively bound by an antibody raised against an antigenic polypeptide, or antigenic fragment thereof, said antigenic polypeptide comprising the polypeptide of any one of SEQ ID NOs: 1-114.

41. The polypeptide of Paragraph 38, wherein said polypeptide is selectively bound by an antibody raised against an antigenic polypeptide, or antigenic fragment thereof, said antigenic polypeptide comprising the polypeptide of any one of SEQ ID NOs: 5, 7, 8 and 11.

42. The polypeptide of Paragraph 38, wherein said polypeptide comprises the polypeptide of SEQ ID NOs: 1-114.

43. The polypeptide of Paragraph 38, wherein said polypeptide comprises a polypeptide selected from the group consisting of SEQ ID NOs. 5, 7, 8 and 11.

44. An antibody that selectively binds to the polypeptide of Paragraph 38.

45. An antibody according to Paragraph 44, wherein said antibody is capable of inhibiting binding of said polypeptide to a THAP-family target polypeptide.

46. An antibody according to Paragraph 44, wherein said antibody is capable of inhibiting apoptosis mediated by said polypeptide.

47. The polyptide of Paragraph 38, wherein identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the default parameters of XBLAST.

48. A method of assessing the biological activity of a THAP-family polypeptide comprising:

(a) providing a THAP-family polypeptide or a fragment thereof; and

(b) assessing the ability of the THAP-family polypeptide to induce apoptosis of a cell.

49. A method of assessing the biological activity of a THAP-family polypeptide comprising:

(a) providing a THAP-family polypeptide or a fragment thereof; and

(b) assessing the DNA binding activity of the THAP-family polypeptide.

50. The method of Paragraphs 48 or 49, wherein step (a) comprises introducing to a cell a recombinant vector comprising a nucleic acid encoding a THAP-family polypeptide.

51. The method of Paragraphs 49 or 50, wherein the THAP-family polypeptide comprises a THAP consensus amino acid sequence depicted in SEQ ID NOs : 1-2, or a fragment thereof having at least one activity selected from the group

consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis.

52. The method of Paragraph 49, wherein the THAP-family polypeptide comprises an amino acid sequence selected from the group of sequences consisting of SEQ ID NOs: 1-114 or a fragment thereof having at least one activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis.

53. The method of Paragraph 49, wherein the THAP-family polypeptide comprises a native THAP-family polypeptide, or a fragment thereof having at least one activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis.

54. The method of Paragraph 49, wherein the THAP-family polypeptide comprises a THAP-family polypeptide or a fragment thereof having at least one activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis, wherein said THAP-family polypeptide or fragment thereof comprises at least one amino acid deletion, substitution or insertion.

55. An isolated THAP-family polypeptide comprising an amino acid sequence of SEQ ID NOs: 1-114, wherein said polypeptide comprises at least one amino acid deletion, substitution or insertion with respect to said amino acid sequence of SEQ ID NOs. 1-114.

56. A THAP-family polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114, wherein said polypeptide comprises at least one amino acid deletion, substitution or insertion with respect to said

amino acid sequence of one of SEQ ID NOs. 1-114 and displays a reduced ability to induce apoptosis or bind DNA compared to the wild-type polypeptide.

57. A THAP-family polypeptide comprising an amino acid sequence of SEQ ID NOs: 1-114, wherein said polypeptide comprises at least one amino acid deletion, substitution or insertion with respect to said amino acid sequence of one of SEQ ID NOs. 1-114 and displays a increased ability to induce apoptosis or bind DNA compared to the wild-type polypeptide.

58. A method of determining whether a THAP-family polypeptide is expressed within a biological sample, said method comprising the steps of :

(a) contacting a biological sample from a subject with:

a polynucleotide that hybridizes under stringent conditions to a nucleic acid of SEQ ID NOs: 160-175 or

a detectable polypeptide that selectively binds to the polypeptide of SEQ ID NOs: 1-114; and

(b) detecting the presence or absence of hybridization between said polynucleotide and an RNA species within said sample, or the presence or absence of binding of said detectable polypeptide to a polypeptide within said sample;

wherein a detection of said hybridization or of said binding indicates that said THAP-family polypeptide is expressed within said sample.

59. The method of Paragraph 58, wherein said subject suffers from, is suspected of suffering from, or is susceptible to a cell proliferative disorder.

60. The method of Paragraph 59, wherein said cell proliferative disorder is a disorder related to regulation of apoptosis.

61. The method of Paragraph 58, wherein said polynucleotide is a primer, and wherein said hybridization is detected by detecting the presence of an amplification product comprising said primer sequence.

62. The method of Paragraph 58, wherein said detectable polypeptide is an antibody.

63. A method of assessing THAP-family activity in a biological sample, said method comprising the steps of :

(a) contacting a nucleic acid molecule comprising a binding site for a THAP-family polypeptide with :

(i) a biological sample from a subject or

(ii) a THAP-family polypeptide isolated from a biological sample from a subject, the polypeptide comprising the amino acid sequences of one of SEQ ID NOs: 1-114; and

(b) assessing the binding between said nucleic acid molecule and a THAP-family polypeptide

wherein a detection of decreased binding compared to a reference THAP-family nucleic acid binding level indicates that said sample comprises a deficiency in THAP-family activity.

64. A method of determining whether a mammal has an elevated or reduced level of THAP-family expression, said method comprising the steps of :

(a) providing a biological sample from said mammal; and

(b) comparing the amount of a THAP-family polypeptide of SEQ ID NOs: 1-114 or of a THAP-family RNA species encoding a polypeptide of SEQ ID NOs: 1-114 within said biological sample with a level detected in or expected from a control sample ;

wherein an increased amount of said THAP-family polypeptide or said THAP-family RNA species within said biological sample compared to said level detected in or expected from said control sample indicates that said mammal has an elevated level of THAP-family expression, and wherein a decreased amount of said THAP-family polypeptide or said THAP-family RNA species within said biological sample compared to said level detected in or expected from said control sample indicates that said mammal has a reduced level oF THAP-family expression.

65. The method of Paragraph 64, wherein said mammal suffers from, is suspected of suffering from, or is susceptible to a cell proliferative disorder.

66. A method of identifying a candidate inhibitor of a THAP-family polypeptide, a candidate inhibitor of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder, said method comprising:

(a) contacting a THAP-family polypeptide according to SEQ ID NOs: 1-114 or a fragment comprising a contiguous span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114 with a test compound; and

(b) determining whether said compound selectively binds to said polypeptide;

wherein a determination that said compound selectively binds to said polypeptide indicates that said compound is a candidate inhibitor of a THAP-family polypeptide, a candidate inhibitor of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder.

67. A method of identifying a candidate inhibitor of apoptosis, a candidate compound for the treatment of a cell proliferative disorder, or a candidate inhibitor of a THAP-family polypeptide of SEQ ID NOs: 1-114 or a fragment comprising a contiguous span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114, said method comprising:

(a) contacting said THAP-family polypeptide with a test compound; and

(b) determining whether said compound selectively inhibits at least one biological activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis;

wherein a determination that said compound selectively inhibits said at least one biological activity of said polypeptide indicates that said compound is a candidate inhibitor of a THAP-family polypeptide, a candidate inhibitor of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder.

68. A method of identifying a candidate inhibitor of apoptosis, a candidate compound for the treatment of a cell proliferative disorder, or a candidate inhibitor of a THAP-family polypeptide of SEQ ID NOs: 1-114 or a fragment comprising a contiguous span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114, said method comprising:

(a) contacting a cell comprising said THAP-family polypeptide with a test compound; and

(b) determining whether said compound selectively inhibits at least one biological activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis;

wherein a determination that said compound selectively inhibits said at least one biological activity of said polypeptide indicates that said compound is a candidate inhibitor of a THAP-family polypeptide, a candidate inhibitor of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder.

69. The method of Paragraphs 67 or 68, wherein step (b) comprises assessing apoptotic activity, and wherein a determination that said compound inhibits apoptosis indicates that said compound is a candidate inhibitor of said THAP-family polypeptide.

70. The method of Paragraph 68 comprising introducing a nucleic acid comprising the nucleotide sequence encoding said THAP-family polypeptide according to any one of Paragraphs 32-43 into said cell.

71. A polynucleotide according to any one of Paragraphs 17- 25 attached to a solid support.

72. An array of polynucleotides comprising at least one polynucleotide according to Paragraph 71.

73. An array according to Paragraph 72, wherein said array is addressable.

74. A polynucleotide according to any one of Paragraphs 17 to 25 further comprising a label.

75. A method of identifying a candidate activator of a THAP-family polypeptide, said method comprising :

a) contacting a THAP-family polypeptide according to SEQ ID NOs: 1-114 or a fragment comprising a a contiguous span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114 with a test compound; and

b) determining whether said compound selectively binds to said polypeptide;

wherein a determination that said compound selectively binds to said polypeptide indicates that said compound is a candidate activator of said polypeptide.

76. A method of identifying a candidate activator of a THAP-family polypeptide of SEQ ID NOs: 1-114 or a fragment comprising a a contiguous span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114, said method comprising:

(a) contacting said polypeptide with a test compound; and

(b) determining whether said compound selectively activates at least one biological activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis;

wherein a determination that said compound selectively activates said at least one biological activity of said polypeptide indicates that said compound is a candidate activator of said polypeptide.

77. A method of identifying a candidate activator of a THAP-family polypeptide of SEQ ID NOs: 1-114 or, a fragment comprising a a contiguous span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114, said method comprising:

(a) contacting a cell comprising said THAP-family polypeptide with a test compound; and

(b) determining whether said compound selectively activates at least one biological activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis;

wherein a determination that said compound selectively activates said at least one biological activity of said polypeptide indicates that said compound is a candidate activator of said polypeptide.

78. The method of Paragraphs 76 or 77, wherein said determining step comprises assessing apoptotic activity, and wherein a determination that said compound increases apoptosis activity indicates that said compound is a candidate activator of said THAP-family polypeptide.

79. The method of Paragraph 77 wherein step a) comprises introducing a nucleic acid comprising the nucleotide sequence encoding said THAP-family polypeptide according to any one of Paragraphs 17-25 into said cell.

80. A method of identifying a candidate modulator of PAR4 activity, said method comprising:

(a) providing a PAR4 polypeptide or a fragment thereof; and

(b) providing a PML-NB polypeptide, or a polypeptide associated with PML-NBs, or a fragment thereof; and

(c) determining whether a test compound selectively modulates the ability of said PAR4 polypeptide to bind to said PML-NB polypeptide or polypeptide associated with PML-NBs;

wherein a determination that said test compound selectively inhibits the ability of said PAR4 polypeptide to bind to said PML-NB polypeptide or polypeptide associated with PML-NBs indicates that said compound is a candidate modulator of PAR4 activity.

81. A method of identifying a candidate modulator of PAR4 activity, said method comprising:

(a) providing a PAR4 polypeptide or a fragment thereof; and

(b) determining whether a test compound selectively modulates the ability of said PAR4 polypeptide to localise in PML-NBs;

wherein a determination that said test compound selectively inhibits the ability of said PAR4 polypeptide to localise in PML-NBs indicates that said compound is a candidate modulator of PAR4 activity.

82. A method of identifying a candidate inhibitor of THAP-family activity, said method comprising:

(a) providing a THAP-family polypeptide of SEQ ID NOs: 1-114 or, a fragment comprising a a contiguous span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114; and

(b) providing a THAP-family target polypeptide or a fragment thereof; and

(c) determining whether a test compound selectively inhibits the ability of said THAP-family polypeptide to bind to said THAP-family target polypeptide;

wherein a determination that said test compound selectively inhibits the ability of said THAP-family polypeptide to bind to said THAP-family target polypeptide indicates that said compound is a candidate inhibitor of THAP-family activity.

83. The method of Paragraph 82, comprising providing a cell comprising:

(a) a first expression vector comprising a nucleic acid encoding a THAP-family polypeptide of SEQ ID NOs: 1-114 or, a fragment comprising a a contiguous span

of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114; and

(b) a second expression vector comprising a nucleic acid encoding a THAP-family target polypeptide, or a fragment thereof.

84. The method of Paragraph 82, wherein said THAP-family activity is apoptosis activity.

85. The method of Paragraph 82, wherein said THAP-family target protein is PAR-4.

86. The method of Paragraph 82, wherein said THAP-family polypeptide is a THAP-1, THAP-2 or THAP-3 protein and said THAP-family target protein is PAR-4.

87. A method of modulating apoptosis in a cell comprising modulating the activity of a THAP-family protein.

88. The method of Paragraph 87, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs. 1-114.

89. A method of modulating apoptosis in a cell comprising modulating the recruitment of PAR-4 to a PML nuclear body.

90. The method of Paragraph 89 wherein modulating the activity of a THAP-family protein comprises modulating the interaction of a THAP-family protein and a THAP-family target protein.

91. The method of Paragraph 89 wherein modulating the activity of a THAP-family protein comprises modulating the interaction of a THAP-family protein and a PAR4 protein.

92. The method of Paragraph 91 comprising modulation the interaction between a THAP-1, THAP-2, or THAP-3 protein and a PAR-4 protein.

93. A method of modulating the recruitment of PAR-4 to a PML nuclear body comprising modulating the interaction of said PAR-4 protein and a THAP-family protein.

94. The method of Paragraph 93, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs. 1-114.

95. A method of modulating angiogenesis in an individual comprising modulating the activity of a THAP-family protein in said individual.

96. The method of Paragraph 95, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs. 1-114.

97. A method of preventing cell death in an individual comprising inhibiting the activity of a THAP-family protein in said individual.

98. The method of Paragraph 97, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs. 1-114.

99. The method according to Paragraph 97, wherein the activity of said THAP-family protein is inhibited in the CNS.

100. A method of inducing angiogenesis in an individual comprising inhibiting the activity of a THAP-family protein in said individual.

101. The method of Paragraph 100, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs. 1-114.

102. A method according to Paragraph 100, wherein the activity of said THAP-family protein is inhibited in endothelial cells.

103. A method of inhibiting angiogenesis or treating cancer in an individual comprising increasing the activity of a THAP-family protein in said individual.

104. The method of Paragraph 103, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs. 1-114.

105. A method of treating inflammation or an inflammatory disorder in an individual comprising increasing the activity of a THAP-family protein in said individual.

106. The method of Paragraph 105, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs. 1-114.

107. A method according to Paragraphs 103 or 105, wherein the activity of said THAP-family protein is increased in endothelial cells.

108. A method of treating cancer in an individual comprising increasing the activity of a THAP-family protein in said individual.

109. The method of Paragraph 108, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs. 1-114.

110. The method of Paragraph 108, wherein increasing the activity of said THAP family protein induces apoptosis, inhibits cell division, inhibits metastatic potential, reduces tumor burden, increases sensitivity to chemotherapy or radiotherapy,

kills a cancer cell, inhibits the growth of a cancer cell, kills an endothelial cell, inhibits the growth of an endothelial cell, inhibits angiogenesis, or induces tumor regression.

111. A method according to any one of Paragraphs 87-110, comprising contacting said subject with a recombinant vector encoding a THAP-family protein according to any one of Paragraphs 32-43 operably linked to a promoter that functions in said cell.

112. The method of Paragraph 111, wherein said promoter functions in an endothelial cell.

113. A viral composition comprising a recombinant viral vector encoding a THAP-family protein according to Paragraphs 32-43.

114. The composition of Paragraph 113, wherein said recombinant viral vector is an adenoviral, adeno-associated viral, retroviral, herpes viral, papilloma viral, or hepatitus B viral vector.

115. A method of obtaining a nucleic acid sequence which is recognized by a THAP-family polypeptide comprising contacting a pool of random nucleic acids with said THAP-family polypeptide or a portion thereof and isolating a complex comprising said THAP-family polypeptide and at least one nucleic acid from said pool.

116. The method of Paragraph 115 wherein said pool of nucleic acids are labeled.

117. The method of Paragraph 116 wherein said complex is isolated by performing a gel shift analysis.

118. A method of identifying a nucleic acid sequence which is recognized by a THAP-family polypeptide comprising:

(a) incubating a THAP-family polypeptide with a pool of labeled random nucleic acids;

(b) isolating a complex between said THAP-family polypeptide and at least one nucleic acid from said pool;

(c) performing an amplification reaction to amplify the at least one nucleic acid present in said complex;

(d) incubating said at least one amplified nucleic acid with said THAP-family polypeptide;

(e) isolating a complex between said at least one amplified nucleic acid and said THAP-family polypeptide;

(f) repeating steps (c), (d) and (e) a plurality of times;

(g) determining the sequence of said nucleic acid in said complex.

119. A method of identifying a compound which inhibits the ability of a THAP-family polypeptide to bind to a nucleic acid comprising :incubating a THAP-family polypeptide or a fragment thereof which recognizes a binding site in a nucleic acid with a nucleic acid containing said binding site in the presence or absence of a test compound and determining whether the level of binding of said THAP-family polypeptide to said nucleic acid in the presence of said test compound is less than the level of binding in the absence of said test compound.

120. A method of identifying a test compound that modulates THAP-mediated activities comprising:

contacting a THAP-family polypeptide or a biologically active fragment thereof with a test compound, wherein said THAP-family polypeptide comprises an amino acid sequence having at least 30% amino acid identity to an amino acid sequence of SEQ ID NO: 1; and

determining whether said test compound selectively modulates the activity of said THAP-family polypeptide or biologically active fragment thereof, wherein a determination that said test compound selectively modulates the activity of said polypeptide indicates that said test compound is a candidate modulator of THAP-mediated activities.

121. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 1, or a biologically active fragment thereof.

122. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 2, or a biologically active fragment thereof.

123. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 3, or a biologically active fragment thereof.

124. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 4, or a biologically active fragment thereof.

125. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 5, or a biologically active fragment thereof.

126. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 6, or a biologically active fragment thereof.

127. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 7, or a biologically active fragment thereof.

128. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 8, or a biologically active fragment thereof.

129. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 9, or a biologically active fragment thereof.

130. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 10, or a biologically active fragment thereof.

131. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 11, or a biologically active fragment thereof.

132. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 12, or a biologically active fragment thereof.

133. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 13, or a biologically active fragment thereof.

134. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence of SEQ ID NO: 14 or a biologically active fragments thereof.

135. The method of Paragraph 120, wherein the THAP-family polypeptide comprises the amino acid sequence selected from the group consisting of SEQ ID NOs: 15-114 or a biologically active fragments thereof.

136. The method of Paragraph 120, wherein said THAP-mediated activity is selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid, binding to PAR-4, binding to SLC, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis

137. The method of Paragraph 136, wherein said THAP-mediated activity is binding to PAR-4.

138. The method of Paragraph 136, wherein said THAP-mediated activity is binding to SLC.

139. The method of Paragraph 136, wherein said THAP-mediated activity is inducing apoptosis.

140. The method of Paragraph 136, wherein said nucleic acid comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 140-159.

141. The method of Paragraph 120, wherein said amino acid identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters, score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the defaul parameters of XBLAST.

142. An isolated or purified THAP domain polypeptide consisting essentially of an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-2, amino acids 1-89 of SEQ ID NOs: 3-5, amino acids 1-90 of SEQ ID NOs: 6-9, amino acids 1-92 of SEQ ID NO: 10, amino acids 1-90 of SEQ ID NOs: 11-14 and homologs having at least 30% amino acid identity to any aforementioned sequence, wherein said polypeptide binds to a nucleic acid.

143. The isolated or purified THAP domain polypeptide of Paragraph 142 consisting essentially of SEQ ID NO: 1.

144. The isolated or purified THAP domain polypeptide of Paragraph 142, wherein said amino acid identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters, score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the defaul parameters of XBLAST.

145. The isolated or purified THAP domain polypeptide of Paragraph 142, wherein said nucleic acid comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 140-159.

146. An isolated or purified nucleic acid which encodes the THAP domain polypeptide of Paragraph 142 or a complement thereof.

147. An isolated or purified PAR4-binding domain polypeptide consisting essentially of an amino acid sequence selected from the group consisting of amino acids 143-192 of SEQ ID NO: 3, amino acids 132-181 of SEQ ID NO: 4, amino acids 186-234 of SEQ ID NO: 5 , SEQ ID NO: 15 and homologs having at least 30% amino acid identity to any aforementioned sequence, wherein said polypeptide binds to PAR4.

148. The isolated or purified PAR4-binding domain of Paragraph 147 consisting essentially of SEQ ID NO: 15.

149. The isolated or purified PAR4-binding domain of Paragraph 147 consisting essentially of amino acids 143-193 of SEQ ID NO: 3.

150. The isolated or purified PAR4-binding domain of Paragraph 147 consisting essentially of amino acids 132-181 of SEQ ID NO: 4.

151. The isolated or purified PAR4-binding domain of Paragraph 147 consisting essentially of amino acids 186-234 of SEQ ID NO: 5.

152. The isolated or purified PAR4-binding domain polypeptide of Paragraph 147, wherein said amino acid identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters, score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the defaul parameters of XBLAST.

153. An isolated or purified nucleic acid which encodes the PAR4-binding domain polypeptide of Paragraph 147 or a complement thereof.

154. An isolated or purified SLC-binding domain polypeptide consisting essentially of an amino acid sequence selected from the group consisting of amino acids

143-213 of SEQ ID NO: 3 and homologs thereof having at least 30% amino acid identity, wherein said polypeptide binds to SLC.

155. The isolated or purified SLC-binding domain polypeptide of Paragraph 154, wherein said amino acid identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters, score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the defaul parameters of XBLAST.

156. An isolated or purified nucleic acid which encodes the SLC-binding domain polypeptide of Paragraph 154 or a complement thereof.

157. A fusion protein comprising an Fc region of an immunoglobulin fused to a polypeptide comprising an amino acid sequence selected from the group consisting of amino acids 143-213 of SEQ ID NO: 3 and homologs thereof having at least 30% amino acid identity.

158. An oligomeric THAP protein comprising a plurality of THAP polypeptides, wherein each THAP polypeptide comprises an amino acid sequence selected from the group consisting of amino acid 143-213 of SEQ ID NO: 3 and homologs thereof having at least 30% amino acid identity.

159. A medicament comprising an effective amount of a THAP1 polypeptide or an SLC-binding fragment thereof, together with a pharmaceutically acceptable carrier.

160. An isolated or purified THAP dimerization domain polypeptide consisting essentially of an amino acid sequence selected from the group consisting of amino acids 143 and 192 of SEQ ID NO: 3 and homologs thereof having at least 30% amino acid identity, wherein said polypeptide binds to a THAP-family polypeptide..

161. The isolated or purified THAP dimerization domain polypeptide of Paragraph 160, wherein said amino acid identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters, score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the defaul parameters of XBLAST.

162. An isolated or purified nucleic acid which encodes the THAP dimerization domain polypeptide of Paragraph 160 or a complement thereof.

163. An expression vector comprising a promoter operably linked to a nucleic acid having a nucleotide sequence selected from the group consisting of SEQ ID NOs: 160-175 and portions thereof comprising at least 18 consecutive nucleotides.

164. The expression vector of Paragraph 163, wherein said promoter is a promoter which is not operably linked to said nucleic acid selected from the group consisting of SEQ ID NOs.: 160-175 in a naturally occurring genome.

165. A host cell comprising the expression vector of Paragraph 163.

166. An expression vector comprising a promoter operably linked to a nucleic acid encoding a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114 and portions thereof comprising at least 18 consecutive nucleotides.

167. The expression vector of Paragraph 166, wherein said promoter is a promoter which is not operably linked to said nucleic acid selected from the group consisting of SEQ ID NOs.: 160-175 in a naturally occurring genome.

168. A host cell comprising the expression vector of Paragraph 166.

169. A method of identifying a candidate inhibitor of a THAP-family polypeptide, a candidate inhibitor of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder, said method comprising:

contacting a THAP-family polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114 or a fragment comprising a span of at least 6 contiguous amino acids of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114 with a test compound; and

determining whether said compound selectively binds to said polypeptide, wherein a determination that said compound selectively binds to said polypeptide indicates that said compound is a candidate inhibitor of a THAP-family polypeptide, a candidate inhibitor of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder.

170. A method of identifying a candidate inhibitor of apoptosis, a candidate compound for the treatment of a cell proliferative disorder, or a candidate inhibitor of a THAP-family polypeptide of SEQ ID NOs: 1-114 or a fragment comprising a span of at

least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114, said method comprising:

contacting said THAP-family polypeptide with a test compound; and

determining whether said compound selectively inhibits at least one biological activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to SLC, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis, wherein a determination that said compound selectively inhibits said at least one biological activity of said polypeptide indicates that said compound is a candidate inhibitor of a THAP-family polypeptide, a candidate inhibitor of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder.

171. A method of identifying a candidate inhibitor of apoptosis, a candidate compound for the treatment of a cell proliferative disorder, or a candidate inhibitor of a THAP-family polypeptide of SEQ ID NOs: 1-114 or a fragment comprising a span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114, said method comprising:

contacting a cell comprising said THAP-family polypeptide with a test compound; and

determining whether said compound selectively inhibits at least one biological activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to SLC, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis, wherein a determination that said compound selectively inhibits said at least one biological activity of said polypeptide indicates that said compound is a candidate inhibitor of a THAP-family polypeptide, a candidate inhibitor of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder.

172. A method of identifying a candidate modulator of THAP-family activity, said method comprising:

providing a THAP-family polypeptide of SEQ ID NOs: 1-114 or, a fragment comprising a span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114; and

providing a THAP-family target polypeptide or a fragment thereof; and

determining whether a test compound selectively modulates the ability of said THAP-family polypeptide to bind to said THAP-family target polypeptide, wherein a determination that said test compound selectively modulates the ability of said THAP-family polypeptide to bind to said THAP-family target polypeptide indicates that said compound is a candidate modulator of THAP-family activity.

173. The method of Paragraph 172, wherein said THAP-family polypeptide is provided by a first expression vector comprising a nucleic acid encoding a THAP-family polypeptide of SEQ ID NOs: 1-114 or, a fragment comprising a contiguous span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-114, and wherein said THAP-family target polypeptide is provided by a second expression vector comprising a nucleic acid encoding a THAP-family target polypeptide, or a fragment thereof.

174. The method of Paragraph 172, wherein said THAP-family activity is apoptosis activity.

175. The method of Paragraph 172, wherein said THAP-family target protein is PAR-4.

176. The method of Paragraph 172, wherein said THAP-family polypeptide is a THAP-1, THAP-2 or THAP-3 protein and said THAP-family target protein is PAR-4.

177. The method of Paragraph 172, wherein said THAP-family target protein is SLC.

178. A method of modulating apoptosis in a cell comprising modulating the activity of a THAP-family protein.

179. The method of Paragraph 178, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs: 1-114.

180. The method of Paragraph 178, wherein modulating the activity of a THAP-family protein comprises modulating the interaction of a THAP-family protein and a THAP-family target protein.

181. The method of Paragraph 178, wherein modulating the activity of a THAP-family protein comprises modulating the interaction of a THAP-family protein and a PAR4 protein.

182. A method of identifying a candidate activator of a THAP-family polypeptide, a candidate activator of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder, said method comprising:

contacting a THAP-family polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-98 or a fragment comprising a span of at least 6 contiguous amino acids of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-98 with a test compound; and

determining whether said compound selectively binds to said polypeptide, wherein a determination that said compound selectively binds to said polypeptide indicates that said compound is a candidate activator of a THAP-family polypeptide, a candidate activator of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder.

183. A method of identifying a candidate activator of apoptosis, a candidate compound for the treatment of a cell proliferative disorder, or a candidate activator of a THAP-family polypeptide of SEQ ID NOs: 1-98 or a fragment comprising a span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-98, said method comprising:

contacting said THAP-family polypeptide with a test compound; and

determining whether said compound selectively activates at least one biological activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to SLC, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis, wherein a determination that said compound selectively activates said at least one biological activity of said polypeptide indicates that said compound is a candidate activator of a THAP-family polypeptide, a candidate activator of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder.

184. A method of identifying a candidate activator of apoptosis, a candidate compound for the treatment of a cell proliferative disorder, or a candidate activator of a THAP-family polypeptide of SEQ ID NOs: 1 to 98 or a fragment comprising a span of at least 6 contiguous amino acids of a polypeptide according to SEQ ID NOs: 1-98, said method comprising:

contacting a cell comprising said THAP-family polypeptide with a test compound; and

determining whether said compound selectively activates at least one biological activity selected from the group consisting of interaction with a THAP-family target protein, binding to a nucleic acid sequence, binding to PAR-4, binding to SLC, binding to PML, binding to a polypeptide found in PML-NBs, localization to PML-NBs, targeting a THAP-family target protein to PML-NBs, and inducing apoptosis, wherein a determination that said compound selectively activates said at least one biological activity of said polypeptide indicates that said compound is a candidate activator of a THAP-family polypeptide, a candidate activator of apoptosis, or a candidate compound for the treatment of a cell proliferative disorder.

185. A method of ameliorating a condition associated with the activity of SLC in an individual comprising administering a polypeptide comprising the SLC binding domain of a THAP-family protein to said individual.

186. The method of Paragraph 185, wherein said polypeptide comprises a fusion protein comprising an Fc region of an immunoglobulin fused to a polypeptide comprising an amino acid sequence selected from the group consisting of amino acids 143-213 of SEQ ID NO: 3 and homologs thereof having at least 30% amino acid identity.

187. The method of Paragraph 185, wherein said polypeptide comprises an oligomeric THAP protein comprising a plurality of THAP polypeptides, wherein each THAP polypeptide comprises an amino acid sequence selected from the group consisting of amino acid 143-213 of SEQ ID NO: 3 and homologs thereof having at least 30% amino acid identity.

188. A method of modulating angiogenesis in an individual comprising modulating the activity of a THAP-family protein in said individual.

189. The method of Paragraph 188, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs: 1-114.

190. The method of Paragraph 188, wherein said modulation is inhibition.

191. The method of Paragraph 188, wherein said modulation is induction.

192. A method of reducing cell death in an individual comprising inhibiting the activity of a THAP-family protein in said individual.

193. The method of Paragraph 192, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs: 1-114.

194. The method according to Paragraph 192, wherein the activity of said THAP-family protein is inhibited in the CNS.

195. A method of reducing inflammation or an inflammatory disorder in an individual comprising modulating the activity of a THAP-family protein in said individual.

196. The method of Paragraph 195, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs: 1-114.

197. A method of reducing the extent of cancer in an individual comprising modulating the activity of a THAP-family protein in said individual.

198. The method of Paragraph 197, wherein said THAP-family protein is selected from the group consisting of SEQ ID NOs: 1-114.

199. The method of Paragraph 197, wherein increasing the activity of said THAP family protein induces apoptosis, inhibits cell division, inhibits metastatic potential, reduces tumor burden, increases sensitivity to chemotherapy or radiotherapy, kills a cancer cell, inhibits the growth of a cancer cell, kills an endothelial cell, inhibits the growth of an endothelial cell, inhibits angiogenesis, or induces tumor regression.

200. A method of forming a complex, said method comprising:

contacting a chemokine with a chemokine-binding agent comprising a polypeptide selected from the group consisting of THAP-1, a polypeptide having at least 30% amino acid identity to THAP-1, a chemokine-binding domain of THAP-1 and a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1, wherein said chemokine and said chemokine binding agent form a complex.

201. The method of Paragraph 200, wherein said amino acid identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters, score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the defaul parameters of XBLAST.

202. The method of Paragraph 200, wherein said polypeptide is fused to an Fc region of an immunoglobulin.

203. The method of Paragraph 200, wherein said polypeptide comprises a THAP dimerization domain.

204. The method of Paragraph 203, wherein said THAP dimerization domain interacts with one or more THAP dimerization domains to form a THAP oligomer.

205. The method of Paragraph 200, wherein said polypeptide is a recombinant polypeptide.

206. The method of Paragraph 200, wherein said chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL9 and CXCL10.

207. The method of Paragraph 200, wherein said chemokine is selected from the group consisting of SLC, CCL19 and CXCL9.

208. The method of Paragraph 200, wherein said polypeptide comprises THAP-1.

209. The method of Paragraph 208, wherein said THAP-1 comprises the amino acid sequence of SEQ ID NO: 3.

210. The method of Paragraph 200, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to THAP-1.

211. The method of Paragraph 200, wherein said polypeptide comprises a chemokine-binding domain of THAP-1.

212. The method of Paragraph 211, wherein said chemokine-binding domain of THAP-1 comprises the amino acid sequence of amino acids 143-213 of SEQ ID NO: 3.

213. The method of Paragraph 200, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

214. A method of inhibiting the activity of a chemokine, said method comprising contacting a chemokine with an effective amount of an agent comprising a

polypeptide selected from the group consisting of THAP-1, a polypeptide having at least 30% amino acid identity to THAP-1, a chemokine-binding domain of THAP-1 and a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1, wherein the activity of said chemokine is inhibited.

215. The method of Paragraph 214, wherein said amino acid identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters, score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the defaul parameters of XBLAST.

216. The method of Paragraph 214, wherein said polypeptide is fused to an Fc region of an immunoglobulin.

217. The method of Paragraph 214, wherein said polypeptide comprises a THAP dimerization domain.

218. The method of Paragraph 217, wherein said THAP dimerization domain interacts with one or more THAP dimerization domains to form a THAP oligomer.

219. The method of Paragraph 214, wherein said polypeptide is a recombinant polypeptide.

220. The method of Paragraph 214, wherein said polypeptide binds to a chemokine selected from the group consisting of SLC, CCL19, CCL5, CXCL9 and CXCL10.

221. The method of Paragraph 214, wherein said polypeptide binds to a chemokine selected from the group consisting of SLC, CCL19 and CXCL9.

222. The method of Paragraph 214, wherein said polypeptide comprises THAP-1.

223. The method of Paragraph 222, wherein said THAP-1 comprises the amino acid sequence of SEQ ID NO: 3.

224. The method of Paragraph 214, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to THAP-1.

225. The method of Paragraph 214, wherein said polypeptide comprises a chemokine-binding domain of THAP-1.

226. The method of Paragraph 225, wherein said chemokine-binding domain of THAP-1 comprises the amino acid sequence of amino acids 143-213 of SEQ ID NO: 3.

227. The method of Paragraph 214, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

228. A method of reducing inflammation comprising administering an effective amount of a chemokine binding agent to a subject afflicted with an inflammatory condition, wherein said chemokine-binding agent comprises a polypeptide selected from the group consisting of THAP-1, a polypeptide having at least 30% amino acid identity to THAP-1, a chemokine-binding domain of THAP-1 and a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

229. The method of Paragraph 228, wherein said amino acid identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters, score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the defaul parameters of XBLAST.

230. The method of Paragraph 228, wherein said polypeptide is fused to an Fc region of an immunoglobulin.

231. The method of Paragraph 228, wherein said polypeptide comprises a THAP dimerization domain.

232. The method of Paragraph 231, wherein said THAP dimerization domain interacts with one or more THAP dimerization domains to form a THAP oligomer.

233. The method of Paragraph 228, wherein said polypeptide is a recombinant polypeptide.

234. The method of Paragraph 228, wherein said polypeptide binds to a chemokine selected from the group consisting of SLC, CCL19, CCL5, CXCL9 and CXCL10.

235. The method of Paragraph 228, wherein said polypeptide binds to a chemokine selected from the group consisting of SLC, CCL19 and CXCL9.

236. The method of Paragraph 228, wherein said polypeptide comprises THAP-1.

237. The method of Paragraph 236, wherein said THAP-1 comprises the amino acid sequence of SEQ ID NO: 3.

238. The method of Paragraph 228, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to THAP-1.

239. The method of Paragraph 228, wherein said polypeptide comprises a chemokine-binding domain of THAP-1.

240. The method of Paragraph 239, wherein said chemokine-binding domain of THAP-1 comprises the amino acid sequence of amino acids 143-213 of SEQ ID NO: 3.

241. The method of Paragraph 228, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

242. A method of reducing one or more symptoms associated with an inflammatory disease, said method comprising administering to a subject afflicted with said inflammatory disease a therapeutically effective amount of an agent which reduces or eliminates the activity of one or more chemokines, wherein said agent comprises a polypeptide selected from the group consisting of THAP-1, a polypeptide having at least 30% amino acid identity to THAP-1, a chemokine-binding domain of THAP-1 and a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

243. The method of Paragraph 242, wherein said polypeptide is fused to an Fc region of an immunoglobulin.

244. The method of Paragraph 242, wherein said polypeptide comprises a THAP dimerization domain.

245. The method of Paragraph 244, wherein said THAP dimerization domain interacts with one or more THAP dimerization domains to form a THAP oligomer.

246. The method of Paragraph 242, wherein said polypeptide is a recombinant polypeptide.

247. The method of Paragraph 242, wherein said polypeptide binds to a chemokine selected from the group consisting of SLC, CCL19, CCL5, CXCL9 and CXCL10.

248. The method of Paragraph 242, wherein said polypeptide binds to a chemokine selected from the group consisting of SLC, CCL19 and CXCL9.

249. The method of Paragraph 242, wherein said polypeptide comprises THAP-1.

250. The method of Paragraph 249, wherein said THAP-1 comprises the amino acid sequence of SEQ ID NO: 3.

251. The method of Paragraph 242, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to THAP-1.

252. The method of Paragraph 242, wherein said polypeptide comprises a chemokine-binding domain of THAP-1.

253. The method of Paragraph 252, wherein said chemokine-binding domain of THAP-1 comprises the amino acid sequence of amino acids 143-213 of SEQ ID NO: 3.

254. The method of Paragraph 242, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

255. The method of Paragraph 242, wherein said inflammatory disease is arthritis.

256. The method of Paragraph 242, wherein said inflammatory disease is inflammatory bowel disease.

257. A method of detecting a chemokine, said method comprising:

contacting a chemokine with a chemokine-binding agent comprising a polypeptide selected from the group consisting of THAP-1, a polypeptide having at least 30% amino acid identity to THAP-1, a chemokine-binding domain of THAP-1 and a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1; and

detecting chemokine-binding agent bound to said chemokine.

258. The method of Paragraph 257, wherein chemokine is selected from the group consisting of SLC, CCL19, CCL5, CXCL9 and CXCL10.

259. The method of Paragraph 257, wherein said chemokine is selected from the group consisting of SLC, CCL19 and CXCL9.

260. A detection system comprising a chemokine-binding agent comprising a polypeptide selected from the group consisting of THAP-1, a polypeptide having at least 30% amino acid identity to THAP-1, a chemokine-binding domain of THAP-1 and

a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1, wherein said chemokine-binding agent is coupled to a solid support.

261. The detection system of Paragraph 260, wherein said polypeptide comprises THAP-1.

262. The detection system of Paragraph 261, wherein said THAP-1 comprises the amino acid sequence of SEQ ID NO: 3.

263. The detection system of Paragraph 260, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to THAP-1.

264. The detection system of Paragraph 260, wherein said polypeptide comprises a chemokine-binding domain of THAP-1.

265. The detection system of Paragraph 264, wherein said chemokine-binding domain of THAP-1 comprises the amino acid sequence of amino acids 143-213 of SEQ ID NO: 3.

266. The detection system of Paragraph 260, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

267. A pharmaceutical composition comprising a chemokine-binding agent in a pharaceutically acceptable carrier, wherein said chemokine-binding agent comprises a polypeptide selected from the group consisting of THAP-1, a polypeptide having at least 30% amino acid identity to THAP-1, a chemokine-binding domain of THAP-1 and a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

268. The pharmaceutical composition of Paragraph 267, wherein said amino acid identity is determined using an algorithm selected from the group consisting of XBLAST with the parameters, score=50 and wordlength=3, Gapped BLAST with the default parameters of XBLAST, and BLAST with the defaul parameters of XBLAST.

269. The pharmaceutical composition of Paragraph 267, wherein said polypeptide is fused to an Fc region of an immunoglobulin.

270. The pharmaceutical composition of Paragraph 267, wherein said polypeptide comprises a THAP dimerization domain.

271. The pharmaceutical composition of Paragraph 270, wherein said THAP dimerization domain interacts with one or more THAP dimerization domains to form a THAP oligomer.

272. The pharmaceutical composition of Paragraph 267, wherein said polypeptide binds to a chemokine selected from the group consisting of SLC, CCL19, CCL5, CXCL9 and CXCL10.

273. The pharmaceutical composition of Paragraph 267, wherein said polypeptide binds to a chemokine selected from the group consisting of SLC, CCL19 and CXCL9.

274. The pharmaceutical composition of Paragraph 267, wherein said polypeptide comprises THAP-1.

275. The pharmaceutical composition of Paragraph 274, wherein said THAP-1 comprises the amino acid sequence of SEQ ID NO: 3.

276. The pharmaceutical composition of Paragraph 267, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to THAP-1.

277. The pharmaceutical composition of Paragraph 267, wherein said polypeptide comprises a chemokine-binding domain of THAP-1.

278. The pharmaceutical composition of Paragraph 277, wherein said chemokine-binding domain of THAP-1 comprises the amino acid sequence of amino acids 143-213 of SEQ ID NO: 3.

279. The pharmaceutical composition of Paragraph 267, wherein said polypeptide comprises a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

280. A device for administering an agent, said device comprising a container that contains therein a chemokine-binding agent in a pharmaceutically acceptable carrier, wherein said chemokine-binding agent comprises a polypeptide selected from the group consisting of THAP-1, a polypeptide having at least 30% amino acid identity to THAP-1, a chemokine-binding domain of THAP-1 and a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1.

281. The device according to Paragraph 280, wherein said container is a syringe.

282. The device according to Paragraph 280, wherein said container is a patch for transdermal administration.

283. The device according to Paragraph 280, wherein said container is pressurized canister.

284. A kit comprising:

a chemokine-binding agent comprising a polypeptide selected from the group consisting of THAP-1, a polypeptide having at least 30% amino acid identity to THAP-1, a chemokine-binding domain of THAP-1 and a polypeptide having at least 30% amino acid identity to a chemokine-binding domain of THAP-1; and

instructions for using said chemokine-binding agent for detecting or inhibiting chemokines.

285. The kit of Paragraph 284, wherein said chemokine is selected from the group consisiting of SLC, CCL19, CCL5, CXCL9 and CXCL10.

286. An isolated or purified chemokine-binding domain consisting essentially of a portion of SEQ ID NO: 3 that binds to a chemokine.

287. The isolated or purified chemokine-binding domain of Paragraph 286, wherein said chemokine is CCL19.

288. The isolated or purified chemokine-binding domain of Paragraph 286, wherein said chemokine is CCL5.

289. The isolated or purified chemokine-binding domain of Paragraph 286, wherein said chemokine is CXCL9.

290. The isolated or purified chemokine-binding domain of Paragraph 286, wherein said chemokine is CXCL10.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A illustrates an amino acid sequence alignment of human THAP1 (hTHAP1) (SEQ ID NO: 3) and mouse THAP1 (mTHAP1) (SEQ ID NO: 99) orthologous polypeptides. Identical amino acid residues are indicated with an asterisk.

Figure 1B depicts the primary structure of the human THAP1 polypeptide. Positions of the THAP domain, the proline-rich region (PRO) and the bipartite nuclear localization sequence (NLS) are indicated.

Figure 2 depicts the results of a Northern Blot analysis of THAP1 mRNA expression in 12 human tissues. Each lane contains 2 μg of poly A^+ RNA isolated from the indicated human tissues. The blot was hybridized, under high-stringency conditions, with a ^{32}P-labeled THAP1 cDNA probe, and exposed at -70°C for 72 hours.

Figure 3A illustrates the interaction between THAP1 and PAR4 in a yeast two-hybrid system. In particular, THAP1 binds to wild-type Par4 (Par4) and the leucine zipper-containing Par4 death domain (Par4DD) (amino acids 250-342 of PAR4) but not a Par4 deletion mutant lacking the death domain (PAR4Δ) (amino acids 1-276 of PAR4). A (+) indicates binding whereas a (-) indicated lack of binding.

Figure 3B shows the binding of *in vitro* translated, ^{35}S-methionine-labeled THAP1 to a GST-Par4DD polypeptide fusion. Par4DD was expressed as a GST fusion protein then purified on an affinity matrix of glutathione sepharose. GST served as negative control. The input represents 1/10 of the material used in the binding assay.

Figure 4A illustrates the interaction between PAR4 and several THAP1 deletion mutants both *in vitro* and *in vivo*. Each THAP1 deletion mutant was tested for binding to either PAR or PAR4DD in a yeast two hybrid system (two hybrid bait), to PAR4DD in GST pull down assays (*in vitro*) and to myc-Par4DD in primary human endothelial cells (*in vivo*). A (+) indicates binding whereas a (-) indicated lack of binding.

Figure 4B shows the binding of several *in vitro* translated, ^{35}S-methionine-labeled THAP1 deletion mutants to a GST-Par4DD polypeptide fusion. Par4DD was expressed as a GST fusion protein then purified on an affinity matrix of glutathione sepharose. GST served as negative control. The input represents 1/10 of the material used in the binding assay.

Figure 5A depicts an amino acid sequence alignment of the Par4 binding domain of human THAP1 (SEQ ID NO: 117) and mouse THAP1 (SEQ ID NO: 116) orthologues with that of mouse ZIP kinase (SEQ ID NO: 115), another Par4 binding partner. An arginine-rich consensus Par4 binding site (SEQ ID NO: 15), derived from this alignment, is also indicated.

Figure 5B shows the primary structure of the THAP1 wild-type polypeptide and two THAP1 mutants (THAP1Δ(QRCRR) and THAP1 RR/AA). THAP1Δ(QRCRR) is a deletion mutant having a deletion of amino acids at positions 168-172 of THAP1

(SEQ ID NO: 3) whereas THAP RR/AA is a mutant having the two arginines located at amino acid positions 171 and 172 to THAP1 (SEQ ID NO: 3) replaced with alanines. Results obtained, in yeast two-hybrid system with Par4 and Par4DD baits (two hybrid bait), in GST pull down assays with GST-Par4DD (*in vitro*) and in the *in vivo* interaction test with myc-Par4DD in primary human endothelial cells (*in vivo*) are summarized.

Figure 6A is a graph which compares apoptosis levels in cells transfected with GFP-APSK1, GFP-Par4 or GFP-THAP1 expression vectors. Apoptosis was quantified by DAPI staining of apoptotic nuclei, 24 h after serum-withdrawal. Values are the means of three independent experiments.

Figure 6B is a graph which compares apoptosis levels in cells transfected with GFP-APSK1 or GFP-THAP1 expression vectors. Apoptosis was quantified by DAPI staining of apoptotic nuclei, 24 h after addition of TNF α. Values are the means of three independent experiments.

Figure 7A shows the binding of *in vitro* translated ^{35}S-methionine labeled THAP1 (wt) or THAP1ΔTHAP (Δ) to a GST-Par4DD polypeptide fusion. Par4DD was expressed as a GST fusion protein then purified on an affinity matrix of glutathione sepharose. GST served as negative control. The input represents 1/10 of the material used in the binding assay.

Figure 7B is a graph which compares the proapoptotic activity of THAP1 with a THAP1 mutant having its THAP domain (amino acids 1-90 of SEQ ID NO: 3) deleted. The percentage of apoptotic cells in mouse 3T3 fibroblasts overexpressing GFP-APSK1 (control), GFP-THAP1 (THAP1) or GFP-THAP1ΔTHAP (THAP1ΔTHAP) was determined by counting apoptotic nuclei after DAPI staining. Values are the means of three independent experiments.

Figure 8 depicts the primary structure of twelve human THAP proteins. The THAP domain (colored grey) is located at the amino-terminus of each of the twelve human THAP proteins. The black box in THAP1, THAP2 and THAP3 indicates a nuclear localization sequence, rich in basic residues, that is conserved in the three proteins. The number of amino-acids in each THAP protein is indicated; (*) indicates the protein is not full length.

Figure 9A depicts an amino acid sequence alignment of the THAP domain of

human THAP1 (hTHAP1, SEQ ID NO: 123) with the DNA binding domain of drosophila melanogaster P-element transposase (dmTransposase, SEQ ID NO: 124). Identical residues are boxed in black and conserved residues in grey. A THAP domain consensus sequence (SEQ ID NO: 125) is also shown.

Figure 9B depicts an amino acid sequence alignment of the THAP domains of twelve members of the human THAP family (hTHAP1, SEQ ID NO: 126; hTHAP2, SEQ ID NO: 131; hTHAP3, SEQ ID NO: 127; hTHAP4, SEQ ID NO: 130; hTHAP5, SEQ ID NO: 128; hTHAP6, SEQ ID NO: 135; hTHAP7, SEQ ID NO: 133; hTHAP8, SEQ ID NO: 129; hTHAP9, SEQ ID NO: 134; hTHAP10, SEQ ID NO: 137; hTHAP11, SEQ ID NO: 136; hTHAP0, SEQ ID NO: 132) with the DNA binding domain of *Drosophila melanogaster* P-element transposase (dmTransposae, SEQ ID NO: 138). Residues conserved among at least seven of the thirteen sequences are boxed. Black boxes indicate identical residues whereas boxes shaded in grey show similar amino acids. Dashed lines represent gaps introduced to align sequences. A THAP domain consensus sequence (SEQ ID NO: 139) is also shown.

Figure 9C depicts an amino acid sequence alignment of 95 distinct THAP domain sequences, including hTHAP1 through hTHAP11 and hTHAP0 (SEQ ID NOs: 3-14, listed sequentially beginning from the top), with 83 THAP domains from other species (SEQ ID NOs: 16-98, listed sequentially beginning at the sequence denoted sTHAP1 and ending at the sequence denoted ceNP_498747.1), which were identified by searching GenBank genomic and EST databases with the human THAP sequences. Residues conserved among at least 50% of the sequences are boxed. Black boxes indicate identical residues whereas boxes shaded in grey show similar amino acids. Dashed lines represent gaps introduced to align sequences. The species are indicated: *Homo sapiens* (h); *Sus scrofa* (s); *Bos taurus* (b); *Mus musculus* (m); *Rattus norvegicus* (r); *Gallus gallus* (g); *Xenopus laevi* (x); *Danio rerio* (z); *Oryzias latipes* (o); *Drosophila melanogaster* (dm); *Anopheles gambiae* (a); *Bombyx mori* (bm); *Caenorhabditis.elegans* (ce). A consensus sequence (SEQ ID NO: 2) is also shown. Amino acids underlined in the consensus sequence are residues which are conserved in all 95 THAP sequences.

Figure 10A shows an amino acid sequence alignment of the human THAP1 (SEQ ID NO: 3), THAP2 (SEQ ID NO: 4) and THAP3 (SEQ ID NO: 5) protein

sequences. Residues conserved among at least two of the three sequences are boxed. Black boxes indicate identical residues whereas boxes shaded in grey show similar amino acids. Dashed lines represent gaps introduced to align sequences. Regions corresponding to the THAP domain, the PAR4-binding domain, and the nuclear localization signal (NLS) are also indicated.

Figure 10B shows the primary structure of human THAP1, THAP2 and THAP3 and results of two-hybrid interactions between each THAP protein and Par4 or Par4 death domain (Par4DD) in the yeast two hybrid system.

Figure 10C shows the binding of *in vitro* translated, ^{35}S-methionine-labeled THAP2 and THAP3 to a GST-Par4DD polypeptide fusion. Par4DD was expressed as a GST fusion protein then purified on an affinity matrix of glutathione sepharose. GST served as negative control. The input represents 1/10 of the material used in the binding assay.

Figure 11A is a graph which compares apoptosis levels in cells transfected with GFP-APSK1, GFP-THAP2 or GFP-THAP3 expression vectors. Apoptosis was quantified by DAPI staining of apoptotic nuclei, 24 h after serum-withdrawal. Values are the means of two independent representative experiments.

Figure 11B is a graph which compares apoptosis levels in cells transfected with GFP-APSK1, GFP-THAP2 or GFP-THAP3 expression vectors. Apoptosis was quantified by DAPI staining of apoptotic nuclei, 24 h after additional of TNFα. Values are the means of two independent representative experiments.

Figure 12 illustrates the results obtained by screening several different THAP1 mutants in a yeast two-hybrid system with SLC/CCL21 bait. The primary structure of each THAP1 deletion mutant that was tested is shown. The 70 carboxy-terminal residues of THAP1 (amino acids 143-213) are sufficient for binding to chemokine SLC/CCL21.

Figure 13 illustrates the interaction of THAP1 with wild type SLC/CCL21 and a SLC/CCL21 mutant deleted of the basic carboxy-terminal extension (SLC/CCL21ΔCOOH). The interaction was analyzed both in yeast two-hybrid system with THAP1 bait and *in vitro* using GST-pull down assays with GST-THAP1.

Figure 14 depicts micrographs of the primary human endothelial cells were transfected with the GFP-THAP0, 1, 2, 3, 6 ,7 ,8 , 10, 11 (green fluorescence)

expression constructs. To reveal the nuclear localization of the human THAP proteins, nuclei were counterstained with DAPI (blue). The bar equals 5 μm.

Figure 15A is a threading–derived structural alignment between the THAP domain of human THAP1 (THAP1) (amino acids 1-81 of SEQ ID NO: 3) and the thyroid receptor β DNA binding domain (NLLB) (SEQ ID NO: 121). The color coding is identical to that described in Figure 15D.

Figure 15B shows a model of the three-dimensional structure of the THAP domain of human THAP1 based on its homology with the crystallographic structure of thyroid receptor β. The color coding is identical to that described in Figure 15D.

Figure 15C shows a model of the three-dimensional structure of the DNA-binding domain of *Drosophila* transposase (DmTRP) based on its homology with the crystallographic structure of the DNA-binding domain of the glucocorticoid receptor. The color coding is identical to that described in Figure 15D.

Figure 15D is a threading–derived structural alignment between the *Drosophila melanogaster* transposase DNA binding domain (DmTRP) (SEQ ID NO: 120) and the glucocorticoid receptor DNA binding domain (GLUA) (SEQ ID NO: 122). In accordance with the sequences and structures in Figures 15A - 15C, the color-coding is the following: brown indicates residues in α-helices; indigo indicates residues in β-strands; red denotes the eight conserved Cys residues in NLLB and GLUA or for the three Cys residues common to THAP1 and DmTRP; magenta indicates other Cys residues in THAP1 or DmTRP; cyan denotes the residues involved in the hydrophobic interactions networks colored in THAP1 or DmTRP.

Figure 16A illustrates the results obtained by screening several different THAP1 mutants in a yeast two-hybrid system with THAP1 bait. The primary structure of each THAP1 deletion mutant that was tested is shown. A (+) indicates binding whereas a (-) indicates no binding.

Figure 16B shows the binding of several *in vitro* translated, ^{35}S-methionine-labeled THAP1 deletion mutants to a GST-THAP1 polypeptide fusion. Wild-type THAP1 was expressed as a GST fusion protein then purified on an affinity matrix of glutathione sepharose. GST served as negative control. The input represents 1/10 of the material used in the binding assay.

Figure 17A is an agarose gel showing two distinct THAP1 cDNA fragments were obtained by RT-PCR. Two distinct THAP1 cDNAs were ~400 and 600 nucleotides in length.

Figure 17B shows that the 400 nucleotide fragment corresponds to an alternatively spliced isoform of human THAP1 cDNA, lacking exon 2 (nucleotides 273-468 of SEQ ID 160).

Figure 17C is a Western blot which shows that the second isoform of human THAP1 (THAP1b) encodes a truncated THAP1 protein (THAP1 C3) lacking the amino-terminal THAP domain.

Figure 18A shows a specific DNA binding site recognized by the THAP domain of human THAP1. The THAP domain recognizes GGGCAA or TGGCAA DNA target sequences preferentially organized as direct repeats with 5 nucleotide spacing (DR-5). The consensus sequence 5'- GGGCAAnnnnnTGGCAA -3' (SEQ ID NO: 149). The DR-5 consensus was generated by examination of 9 nucleic acids bound by THAP1 (SEQ ID NO: 140-148, beginning sequentially from the top).

Figure 18B shows a second specific DNA binding site recognized by the THAP domain of human THAP1. The THAP domain recognizes everted repeats with 11 nucleotide spacing (ER-11) having a consensus sequence 5'-TTGCCAnnnnnnnnnnnGGGCAA -3' (SEQ ID NO: 159). The ER-11 consensus was generated by examination of 9 nucleic acids bound by THAP1 (SEQ ID NO: 150-158, beginning sequentially from the top).

Figure 19 shows that THAP1 interacts with both CC and CXC chemokines both *in vivo* in a yeast two-hybrid system with THAP1 prey and *in vitro* using GST-pull down assays with immobilized GST-THAP1. The cytokine IFNγ was used as a negative control. Results are summarized as follows: +++ indicates strong binding; ++ indicates intermediate binding; +/- indicates some binding; - indicates no binding; and ND indicates not determined.

Figure 20A is an SDS-polyacrylamide gel showing the relative amounts of chemokine and cytokine used in immobilized GST-THAP1 binding assays.

Figure 20B is an SDS-polyacrylamide gel showing that neither the cytokine, IFNγ, nor any of the chemokines bound to immobilized GST alone.

Figure 20c is an SDS-polyacrylamide gel showing that chemokines, CXCL10, CXCL9 and CCL19, but not the cytokine IFNγ, bound to immobilized GST-THAP1 fusions.

DETAILED DESCRIPTION OF THE INVENTION

THAP and PAR4 biological pathways

As mentioned above, the inventors have discovered a novel class of proteins involved in apoptosis. Then, the inventors have also linked a member of this novel class to another (PAR4) apoptosis pathway, and further linked both of these pathways to PML-NBs. Moreover, the inventors have also linked both of these pathways to endothelial cells, providing a range of novel and potentially selective therapeutic treatments. In particular, it has been discovered that THAP1 (THanatos (death)-Associated-Protein-1) localizes to PML-NBs. Furthermore, two hybrid screening of an HEVEC cDNA library with the THAP1 bait lead to the identification of a unique interacting partner, the pro-apoptotic protein PAR4. PAR4 is also found to accumulate into PML-NBs. Targeting of the THAP-1 / PAR4 complex to PML-NBs is mediated by PML. Similarly to PAR4, THAP1 has a pro-apoptotic activity. This activity includes a novel motif in the amino-terminal part called THAP domain. Together these results define a novel PML-NBs pathway for apoptosis that involves the THAP1/PAR4 pro-apoptotic complex.

THAP-family members, and uses thereof

The present invention includes polynucleotides encoding a family of pro-apoptotic polypeptides THAP-0 to THAP11, and uses thereof for the modulation of apoptosis-related and other THAP-mediated activities. Included is THAP1, which forms a complex with the pro-apoptotic protein PAR4 and localizes in discrete subnuclear domains known as PML nuclear bodies. Additionally, THAP-family polypeptides can be used to alter or otherwise modulate bioavailability of SLC/CCL21 (SLC).

The present invention also includes a novel protein motif, the THAP domain, which is found in an 89 amino acid domain in the amino-terminal part of THAP1 and

which is involved in THAP1 pro-apoptotic activity. The THAP domain defines a novel family of proteins, the THAP-family, with at least twelve distinct members in the human genome (THAP-0 to THAP11), which all contain a THAP domain in their amino-terminal part. The present invention thus pertains to nucleic acid molecules, including genomic and in particular the complete cDNA sequences, encoding members of the THAP-family, as well as with the corresponding translation products, nucleic acids encoding THAP domains, homologues thereof, nucleic acids encoding at least 10, 12, 15, 20, 25, 30, 40, 50, 100,150 or 200 consecutive amino acids, to the extent that said span is consistent with the particular SEQ ID NO, of a sequence selected from the group consisting of SEQ ID NOs: 160-175.

THAP1 has been identified based on its expression in HEVs, specialized postcapillary venules found in lymphoid tissues and nonlymphoid tissues during chronic inflammatory diseases that support a high level of lymphocyte extravasation from the blood. An important element in the cloning of the THAP1 cDNA from HEVECs was the development of protocols for obtaining HEVECs RNA, since HEVECs are not capable of maintaining their phenotype outside of their native environment for more than a few hours. A protocol was developed where total RNA was obtained from HEVECs freshly purified from human tonsils. Highly purified HEVECs were obtained by a combination of mechanical and enzymatic procedures, immunomagnetic depletion and positive selection. Tonsils were minced finely with scissors on a steel screen, digested with collagenase/dispase enzyme mix and unwanted contaminating cells were then depleted using immunomagnetic depletion. HEVECs were then selected by immunomagnetic positive selection with magnetic beads conjugated to the HEV-specific antibody MECA-79. From these HEVEC that were 98% MECA-79-positive, 1 μg of total RNA was used to generate full length cDNAs for THAP1 cDNA cloning and RT-PCR analysis.

As used herein, the term "nucleic acids" and "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. Throughout the present specification, the expression "nucleotide sequence" may be employed to designate indifferently a polynucleotide or a nucleic

acid. More precisely, the expression "nucleotide sequence" encompasses the nucleic material itself and is thus not restricted to the sequence information (i.e. the succession of letters chosen among the four base letters) that biochemically characterizes a specific DNA or RNA molecule. Also, used interchangeably herein are terms "nucleic acids", "oligonucleotides", and "polynucleotides".

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated THAP-family nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having the nucleotide sequence of SEQ ID NOs: 160-175, a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NOs: 160-175, as a hybridization probe, THAP-family nucleic acid molecules can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. Molecular Cloning. A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989).

Moreover, a nucleic acid molecule encompassing all or a portion of e.g. SEQ ID NOs: 160-175, can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NOs: 160-175.

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis.

Furthermore, oligonucleotides corresponding to THAP-family nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

As used herein, the term "hybridizes to" is intended to describe conditions for moderate stringency or high stringency hybridization, preferably where the hybridization and washing conditions permit nucleotide sequences at least 60% homologous to each other to remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85%, 90%, 95% or 98% homologous to each other typically remain hybridized to each other. Stringent conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are as follows: the hybridization step is realized at 65°C in the presence of 6 x SSC buffer, 5 x Denhardt's solution, 0,5% SDS and 100µg/ml of salmon sperm DNA. The hybridization step is followed by four washing steps:

- two washings during 5 min, preferably at 65°C in a 2 x SSC and 0.1%SDS buffer;

- one washing during 30 min, preferably at 65°C in a 2 x SSC and 0.1% SDS buffer,

- one washing during 10 min, preferably at 65°C in a 0.1 x SSC and 0.1%SDS buffer,

these hybridization conditions being suitable for a nucleic acid molecule of about 20 nucleotides in length. It will be appreciated that the hybridization conditions described above are to be adapted according to the length of the desired nucleic acid, following techniques well known to the one skilled in the art, for example be adapted according to the teachings disclosed in Hames B.D. and Higgins S.J. (1985) *Nucleic Acid Hybridization: A Practical Approach.* Hames and Higgins Ed., IRL Press, Oxford; and Current Protocols in Molecular Biolog (supra). Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence of SEQ ID NOs: 160-175 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA

molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, 90% or 95% of the length of the reference sequence (e.g., when aligning a second sequence to e.g. a THAP-1 amino acid sequence of SEQ ID NO: 3 having 213 amino acid residues, at least 50, preferably at least 100, more preferably at least 200, amino acid residues are aligned or when aligning a second sequence to the THAP-1 cDNA sequence of SEQ ID NO: 160 having 2173 nucleotides or nucleotides 202- 835 which encode the amino acids of the THAP1 protein, preferably at least 100, preferably at least 200, more preferably at least 300, even more preferably at least 400, and even more preferably at least 500, 600, at least 700, at least 800, at least 900, at least 1000, at least 1200, at least 1400, at least 1600, at least 1800, or at least 2000 nucleotides are aligned. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % homology = number (#) of identical positions/total number (#) of positions 100).

The comparison of sequences and determination of percent homology between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Karlin and Altschul (1990) Proc. Natl. Acad. Sci. USA 87:2264-68, modified as in Karlin and Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-77, the

disclosures of which are incorporated herein by reference in their entireties. Such an algorithm is incorporated into the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) J. Mol. Biol. 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score=100, wordlength=12 to obtain nucleotide sequences homologous to THAP-family nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score=50, wordlength=3 to obtain amino acid sequences homologous to THAP-family protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) Nucleic Acids Research 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used (see, www.ncbi.nlm.nih.gov, the disclosures of which are incorporated herein by reference in their entireties). Another preferred, non-limiting example of a mathematical algorithim utilized for the comparison of sequences is the algorithm of Myers and Miller, CABIOS (1989), the disclosures of which are incorporated herein by reference in their entireties. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

The term "polypeptide" refers to a polymer of amino acids without regard to the length of the polymer; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not specify or exclude post-expression modifications of polypeptides, for example, polypeptides which include the covalent attachment of glycosyl groups, acetyl groups, phosphate groups, lipid groups and the like are expressly encompassed by the term polypeptide. Also included within the definition are polypeptides which contain one or more analogs of an amino acid (including, for example, non-naturally occurring amino acids, amino acids which only occur naturally in an unrelated biological system, modified amino acids from mammalian systems etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or

tissue source from which the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of a protein according to the invention (e.g. THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof) in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of a protein according to the invention having less than about 30% (by dry weight) of protein other than the THAP-family protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of protein other than the protein according to the invention, still more preferably less than about 10% of protein other than the protein according to the invention, and most preferably less than about 5% of protein other than the protein according to the invention. When the protein according to the invention or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of a THAP-family protein having less than about 30% (by dry weight) of chemical precursors or non-THAP-family chemicals, more preferably less than about 20% chemical precursors or non-THAP-family or THAP-domain chemicals, still more preferably less than about 10% chemical precursors or non-THAP-family or THAP-domain chemicals, and most preferably less than about 5% chemical precursors or non-THAP-family or THAP-domain chemicals.

The term "recombinant polypeptide" is used herein to refer to polypeptides that have been artificially designed and which comprise at least two polypeptide sequences that are not found as contiguous polypeptide sequences in their initial natural

environment, or to refer to polypeptides which have been expressed from a recombinant polynucleotide.

Accordingly, another aspect of the invention pertains to anti-THAP-family or THAP-domain antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of a THAP-family or THAP domain polypeptide. A monoclonal antibody composition thus typically displays a single binding affinity for a particular THAP-family or THAP domain protein with which it immunoreacts.

PAR4

As mentioned above, Prostate apoptosis response-4 (PAR4) is a 38 kDa protein initially identified as the product of a gene specifically upregulated in prostate tumor cells undergoing apoptosis (for reviews see Rangnekar, 1998 ; Mattson et al., 1999). The PAR4 nucleic acid and amino acid sequences, see Johnstone et al, Mol. Cell. Biol. 16 (12), 6945-6956 (1996); and Genbank accession no. U63809 (SEQ ID NO: 118).

As used interchangeably herein, a "PAR4 activity", "biological activity of a PAR4" or "functional activity of a PAR4", refers to an activity exerted by a PAR4 protein, polypeptide or nucleic acid molecule as determined in vivo, or in vitro, according to standard techniques. In one embodiment, a PAR4 activity is a direct activity, such as an association with a PAR4-target molecule or most preferably apoptosis induction activity, or inhibition of cell proliferation or cell cycle. As used herein, a "target molecule" is a molecule with which a PAR4 protein binds or interacts

in nature, such that PAR4-mediated function is achieved. An example of a PAR4 target molecule is a THAP-family protein such as THAP1 or THAP2, or a PML-NBs protein. A PAR4 target molecule can be a PAR4 protein or polypeptide or a non-PAR4 molecule. For example, a PAR4 target molecule can be a non-PAR4 protein molecule. Alternatively, a PAR4 activity is an indirect activity, such as an activity mediated by interaction of the PAR4 protein with a PAR4 target molecule such that the target molecule modulates a downstream cellular activity (e.g., interaction of a PAR4 molecule with a PAR4 target molecule can modulate the activity of that target molecule on an intracellular signaling pathway).

Binding or interaction with a PAR4 target molecule (such as THAP1/PAR4 described herein) or with other targets can be detected for example using a two hybrid-based assay in yeast to find drugs that disrupt interaction of the PAR4 bait with the target (e.g. PAR4) prey, or an in vitro interaction assay with recombinant PAR4 and target proteins (e.g. THAP1 and PAR4).

CHEMOKINES

Chemokines (chemoattractant cytokines) are small secreted polypeptides of about 70-110 amino acids that regulate trafficking and effector functions of leukocytes, and play an important role in inflammation and host defence against pathogens (reviewed in Baggiolini M., et al. (1997) Annu. Rev. inmmunol. 15: 675-705; Proost P., et al. (1996) Int. J. Clin. Lab. Rse. 26: 211-223; Premack, et al. (1996) Nature Medicine 2: 1174-1178; Yoshie, et al. (1997) J. Leukocyte Biol. 62: 634-644). Over 45 different human chemokines have been described to date. They vary in their specificities for different leukocyte types (neutrophils, monocytes, eosinophils, basophils, lymphocytes, dendritic cells, etc.), and in the types of cells and tissues where the chemokines are synthesized. Chemokines are typically produced at sites of tissue injury or stress, where they promote the infiltration of leukocytes into tissues and facilitate an inflammatory response. Some chemokines act selectively on immune system cells such as subsets of T-cells or B lymphocytes or antigen presenting cells, and may thereby promote immune responses to antigens. Some chemokines also have the ability to regulate the growth or migration of hematopoietic progenitor and stem cells that normally differentiate into specific leukocyte types, thereby regulating leukocyte numbers in the blood.

The activities of chemokines are mediated by cell surface receptors which are members of the family of seven transmembrane, G-protein coupled receptors. At present, over fifteen different human chemokine receptors are known, including CCR1, CCR2, CCR3, CCR4, CCR5, CCR6, CCR7, CCR8, CCR9, CCR10, CXCR1, CXCR2, CXCR3, CXCR4 and CXCR5. These receptors vary in their specificites for specific chemokines. Some receptors bind to a single known chemokine, while others bind to multiple chemokines. Binding of a chemokine to its receptor typically induces intracellular signaling responses such as a transient rise in cytosolic calcium concentration, followed by cellular biological responses such as chemotaxis.

Chemokines are important in medicine because they regulate the movement and biological activities of leukocytes in many disease situations, including, but not limited to: allergic disorders, autoimmune diseases, ischemia/reperfusion injury, development of atherosclerotic plaques, cancer (including mobilization of hematopoietic stem cells for use in chemotherapy or myeloprotection during chemotherapy), chronic inflammatory disorders, chronic rejection of transplanted organs or tissue grafts, chronic myelogenous leukemia, and infection by HIV and other pathogens. Antagonists of chemokines or chemokine receptors may be of benefit in many of these diseases by reducing excessive inflammation and immune system responses.

The activity of chemokines is tightly regulated to prevent excessive inflammation that can cause disease. Inhibition of chemokines by neutralizing antibodies in animal models (Sekido et al. (1993) Nature 365:654-657) or disruption of mouse chemokine genes (Cook et al. (1995) Science 269:1583-1588) have confirmed a critical role of chemokines *in vivo* in inflammation mediated by virus infection or other processes. The production of soluble versions of cytokine receptors containing only the extracellular binding domain, represents a physiological and therapeutic strategy to block the activity of some cytokines (Rose-John and Heinrich (1994) Biochem J. 300:281-290; Heaney and Golde (1996) Blood 87:847-857). However, the seven transmembrane domain structure of chemokine receptors makes the construction of soluble, inhibitory receptors difficult, and thus antagonists based on mutated chemokines, blocking peptides or antibodies are under evaluation as chemokine inhibitors (D'Souza & Harden (1996) Nature Medecine 2:1293-1300; Howard et al.

(1996) Trends Biotech. 14:46-51; Baggiolini (1998) Nature 392:565-568; Rollins (1997) Blood 90:909-928).

Several viral chemokine binding proteins have been described that may be useful as soluble chemokine inhibitors. Soluble chemokine-binding proteins have been previously detected in poxviruses. Firstly, the myxoma virus T7 protein, which was first identified as a soluble IFN-γ Receptor (Upton et al. (1992) Science 258:1369-1372), binds to a range of chemokines through the heparin-binding domain and affects the infiltration of cells into infected tissue (Lalani et al. (1997) J Virol 71:4356-4363). The protein is described in U.S. Patent No. 5,834,419 and International Publication No. WO 96/33730, and is designated CBP-1. Secondly, it was demonstrated that VV strain Lister expresses a soluble 35 kDa protein that is secreted from infected cells and which binds many CC chemokines (Graham et al. (1997) Virology 229:12-24; Smith et al. (1997) Virology 236:316-327; Alcami et al (1998) J Immunol 160:624-633), but not CXC chemokines, through a domain distinct from the heparin-binding domain (Smith et al. (1997) Virology 236:316-327; Alcami et al (1998) J Immunol 160:624-633). This protein has been called vCKBP (Alcami et al (1998) J Immunol 160:624-633). The protein is also described in U.S. Patent No. 5,871,740 and International Publication No. WO97/11714. One main disadvantage to the use of these viral proteins in a clinical setting is that antigenicity severely limits their indications. As such, there is a strong interest in the identification of cellular chemokine-binding proteins.

Some aspects of the present invention relate to cellular polypeptides and homologs thereof, portions of cellular polypeptides and homologs thereof as well as modified cellular polypeptides and homologs thereof that bind to one or more chemokines. In some embodiments of the present invention such cellular polypeptides are THAP-family polypeptides, including THAP-1, chemokine-binding domains of THAP-family polypeptides (including a chemokine-binding domain of THAP-1), THAP-family polypeptide or THAP-family chemokine-binding domain fusions to immunoglobulin Fc (including THAP-1 fused to an immunoglobulin Fc region or a chemokine-binding domain of THAP-1 fused to an immunoglobulin Fc region), oligomers of THAP-family polypeptides or THAP-family chemokine-binding domains (including THAP-1 oligomers or oligomers of a chemokine-binding domain of THAP-1), or homologs of any of the above-listed compositions. Throughout this disclosure,

the above-listed polypeptides are referred to as THAP-type chemokine-binding agents. Each of these THAP-type chemokine-binding agents are described in detail below.

SLC/CCL21 (SLC)

Biological Roles of SLC

The signals which mediate T-cell infiltration during T-cell auto-immune diseases are poorly understood. SLC/CCL21 (SEQ ID NO: 119) is highly potent and highly specific for attracting T-cell migration. It was initially thought to be expressed only in secondary lymphoid organs, directing naive T-cells to areas of antigen presentation. However, using immunohistology it was found that expression of CCL21 was highly induced in endothelial cells of T-cell auto-immune infiltrative skin diseases (Christopherson et al. (2002) Blood electronic publication prior to printed publication). No other T-cell chemokine was consistently induced in these T cell skin diseases. The receptor for CCL21, CCR7, was also found to be highly expressed on the infiltrating T-cells, the majority of which expressed the memory CD45Ro phenotype. Inflamed venules endothelial cells expressing SLC/CCL21 in T cell infiltrative autoimmune skin diseases may therefore play a key role in the regulation of T-cell migration into these tissues.

There are a number of other autoimmune diseases where induced expression of SLC/CCL21 in endothelial cells may cause abnormal recruitment of T-cells from the circulation to sites of pathologic inflammation. For instance, chemokine SLC/CCL21 appears to be important for aberrant T-cell infiltration in experimental autoimmune encephalomyelitis (EAE), an animal model for multiple sclerosis (Alt et al. (2002) Eur J Immunol 32:2133-44). Migration of autoaggressive T cells across the blood-brain barrier (BBB) is critically involved in the initiation of EAE. The direct involvement of chemokines in this process was suggested by the observation that G-protein-mediated signaling is required to promote adhesion strengthening of encephalitogenic T cells on BBB endothelium in vivo. A search for chemokines present at the BBB, by in situ hybridizations and immunohistochemistry revealed expression of the lymphoid chemokines CCL19/ELC and CCL21/SLC in venules surrounded by inflammatory cells (Alt et al. (2002) Eur J Immunol 32:2133-44). Their expression was paralleled by the presence of their common receptor CCR7 in inflammatory cells in brain and spinal cord

sections of mice afflicted with EAE. Encephalitogenic T cells showed surface expression of CCR7 and specifically chemotaxed towards both CCL19 or CCL21 in a concentration dependent and pertussis toxin-sensitive manner comparable to naive lymphocytes in vitro. Binding assays on frozen sections of EAE brains demonstrated a functional involvement of CCL19 and CCL21 in adhesion strengthening of encephalitogenic T lymphocytes to inflamed venules in the brain (Alt et al. (2002) Eur J Immunol 32:2133-44). Taken together these data suggested that the lymphoid chemokines CCL19 and CCL21 besides regulating lymphocyte homing to secondary lymphoid tissue are involved in T lymphocyte migration into the immunoprivileged central nervous system during immunosurveillance and chronic inflammation.

Other diseases where induced expression of SLC/CCL21 in venular endothelial cells has been observed include rheumatoid arthritis (Page et al. (2002) J Immunol 168:5333-5341) and experimental autoimmune diabetes (Hjelmstrom et al. (2000) Am J Path 156:1133-1138). Therefore, chemokine SLC/CCL21 may be an important pharmacological target in T-cell auto-immune diseases. Inhibitors of SLC/CCL21 may be effective agents at treating these T cell infiltrative diseases by interfering with the abnormal recruitment of T cells, from the circulation to sites of pathologic inflammation, by endothelial cells expressing SLC/CCL21. The reduction in T cell migration into involved tissue would reduce the T-cell inflicted damage seen in those diseases.

Ectopic lymphoid tissue formation is a feature of many chronic inflammatory diseases, including rheumatoid arhtritis, inflammatory bowel diseases (Crohn's disease, ulcerative colitis), autoimmune diabetes, chronic inflammatory skin diseases (lichen panus, psoriasis, ...), Hashimoto's thyroiditis, Sjogren's syndrome, gastric lymphomas and chronic inflammatory liver disease (Girard and Springer (1995) Immunol today 16:449-457; Takemura et al. (2001) J Immunol 167:1072-1080; Grant et al. (2002) Am J Pathol 2002 160:1445-55; Yoneyama et al. (2001) J Exp Med 193:35-49).

Ectopic expression of SLC/CCL21 has been shown to induce lymphoid neogenesis, both in mice and in human inflammatory diseases. In mice, transgenic expression of SLC/CCL21 in the pancreas (Fan et al. (2000) J Immunol 164:3955-3959; Chen et al. (2002) J Immunol 168:1001-1008; Luther et al. (2002) J Immunol 169:424-433), a non-lymphoid tissue, has been found to be sufficient for the development and

organization of ectopic lymphoid tissue through differential recruitment of T and B lymphocytes and induction of high endothelial venules, specialized blood vessels for lymphocyte migration (Girard and Springer (1995) Immunol today 16:449-457). In humans, hepatic expression of SLC/CCL21 has been shown to promote the development of high endothelial venules and portal-associated lymphoid tissue in chronic inflammatory liver disease (Grant et al. (2002) Am J Pathol 2002 160:1445-55; Yoneyama et al. (2001) J Exp Med 193:35-49).The chronic inflammatory liver disease primary sclerosing cholangitis (PSC) is associated with portal inflammation and the development of neolymphoid tissue in the liver. More than 70% of patients with PSC have a history of inflammatory bowel disease and strong induction of SLC/CCL21 on CD34(+) vascular endothelium in portal associated lymphoid tissue in PSC has been reported (Grant et al. (2002) Am J Pathol 2002 160:1445-55). In contrast, CCL21 is absent from LYVE-1(+) lymphatic vessel endothelium. Intrahepatic lymphocytes in PSC include a population of CCR7(+) T cells only half of which express CD45RA and which respond to CCL21 in migration assays. The expression of CCL21 in association with mucosal addressin cell adhesion molecule-1 in portal tracts in PSC may promote the recruitment and retention of CCR7(+) mucosal lymphocytes leading to the establishment of chronic portal inflammation and the expanded portal-associated lymphoid tissue. These findings are supported by studies in an animal model of chronic hepatic inflammation, that have shown that anti-SLC/CCL21 antibodies prevent the development of high endothelial venules and portal-associated lymphoid tissue (Yoneyama et al. (2001) J Exp Med 193:35-49).

Induction of chemokine SLC/CCL21 at a site of inflammation could convert the lesion from an acute to a chronic state with corresponding development of ectopic lymphoid tissue. Blocking chemokine SLC/CCL21 activity in chronic inflammatory diseases may therefore have significant therapeutic value.

As used herein, "SLC/CCL21" and "SLC" are synonymous.

THAP-family members comprising a THAP Domain

Based on the elucidation of a biological activity of the THAP1 protein in apoptosis as described herein, the inventors have identified and further characterized a novel protein motif, referred to herein as THAP domain. The THAP domain has been

identified by the present inventors in several other polypeptides, as further described herein. Knowledge of the structure and function of the THAP domain allows the performing of screening assays that can be used in the preparation or screening of medicaments capable of modulating interaction with a THAP-family-target molecule, modulating cell cycle and cell proliferation, inducing apoptosis or enhancing or participating in the induction of apoptosis.

As used interchangeably herein, a THAP-family protein or polypeptide, or a THAP-family member refers to any polypeptide having a THAP domain as described herein. As mentioned, the inventors have provided several specific THAP-family members. Thus, as referred to herein, a THAP-family protein or polypeptide, or a THAP-family member, includes but is not limited to a THAP-0, THAP1, THAP-2, THAP-3, THAP-4, THAP-5, THAP-6, THAP-7, THAP-8, THAP-9, THAP10 or a THAP11 polypeptide.

As used interchangeably herein, a "THAP-family activity", "biological activity of a THAP-family member" or "functional activity of a THAP-family member", refers to an activity exerted by a THAP family or THAP domain polypeptide or nucleic acid molecule, or a biologically active fragment or homologue thereof comprising a THAP as determined in vivo, or in vitro, according to standard techniques. In one embodiment, a THAP-family activity is a direct activity, such as an association with a THAP-family-target molecule or most preferably apoptosis induction activity, or inhibition of cell proliferation or cell cycle. As used herein, a " THAP-family target molecule" is a molecule with which a THAP-family protein binds or interacts in nature, such that a THAP family-mediated function is achieved. For example, a THAP family target molecule can be another THAP-family protein or polypeptide which is substantially identical or which shares structural similarity (e.g. forming a dimer or multimer). In another example, a THAP family target molecule can be a non-THAP family comprising protein molecule, or a non-self molecule such as for example a Death Domain receptor. Binding or interaction with a THAP family target molecule (such as THAP1/PAR4 described herein) or with other targets can be detected for example using a two hybrid-based assay in yeast to find drugs that disrupt interaction of the THAP family bait with the target (e.g. PAR4) prey, or an in vitro interaction assay with recombinant THAP family and target proteins (e.g. THAP1 and PAR4). In yet another

example, a THAP family target molecule can be a nucleic acid molecule. For instance, a THAP family target molecule can be DNA.

Alternatively, a THAP-family activity may be an indirect activity, such as an activity mediated by interaction of the THAP-family protein with a THAP-family target molecule such that the target molecule modulates a downstream cellular activity (e.g., interaction of a THAP-family molecule with a THAP-family target molecule can modulate the activity of that target molecule on an intracellular signaling pathway).

THAP-family activity is not limited to the induction of apoptotic activity, but may also involve enhancing apoptotic activity. As death domains may mediate protein-protein interactions, including interactions with other death domains, THAP-family activity may involve transducing a cytocidal signal.

Assays to detect apoptosis are well known. In a preferred example, an assay is based on serum-withdrawal induced apoptosis in a 3T3 cell line with tetracycline-regulated expression of a THAP family member comprising a THAP domain. Other non-limiting examples are also described.

In one example, a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof can be the minimum region of a polypeptide that is necessary and sufficient for the generation of cytotoxic death signals. Exemplary assays for apoptosis activity are further provided herein.

In specific embodiments, PAR4 is a preferred THAP1 and/or THAP2 target molecule. In another aspect, a THAP1 target molecule is a PML-NB protein.

In further aspects, THAP-domain or a THAP-family polypeptide comprises a DNA binding domain.

In other aspects, a THAP-family activity is detected by assessing any of the following activities: (1) mediating apoptosis or cell proliferation when expressed in or introduced into a cell, most preferably inducing or enhancing apoptosis, and/or most preferably reducing cell proliferation; (2) mediating apoptosis or cell proliferation of an endothelial cell; (3) mediating apoptosis or cell proliferation of a hyperproliferative cell; (4) mediating apoptosis or cell proliferation of a CNS cell, preferably a neuronal or glial cell; (5) an activity determined in an animal selected from the group consisting of mediating, preferably inhibiting angiogenesis, mediating, preferably inhibiting inflammation, inhibition of metastatic potential of cancerous tissue, reduction of tumor

burden, increase in sensitivity to chemotherapy or radiotherapy, killing a cancer cell, inhibition of the growth of a cancer cell, or induction of tumor regression; or (6) interaction with a THAP family target molecule or THAP domain target molecule, preferably interaction with a protein or a nucleic acid. Detecting THAP-family activity may also comprise detecting any suitable therapeutic endpoint discussed herein in the section titled "Methods of Treatment". THAP-family activity may be assessed either in vitro (cell or non-cell based) or in vivo depending on the assay type and format.

A THAP domain has been identified in the N-terminal region of the THAP1 protein, from about amino acid 1 to about amino acid 89 of SEQ ID NO: 3 based on sequence analysis and functional assays. A THAP domain has also been identified in THAP2 to THAP0 of SEQ ID NOs: 4-14. However, it will be appreciated that a functional THAP domain may be only a small portion of the protein, about 10 amino acids to about 15 amino acids, or from about 20 amino acids to about 25 amino acids, or from about 30 amino acids to about 35 amino acids, or from about 40 amino acids to about 45 amino acids, or from about 50 amino acids to about 55 amino acids, or from about 60 amino acids to about 70 amino acids, or from about 80 amino acids to about 90 amino acids, or about 100 amino acids in length. Alternatively, THAP domain or THAP family polypeptide activity, as defined above, may require a larger portion of the native protein than may be defined by protein-protein interaction, DNA binding, cell assays or by sequence alignment. A portion of a THAP domain-containing polypeptide from about 110 amino acids to about 115 amino acids, or from about 120 amino acids to 130 amino acids, or from about 140 amino acids to about 150 amino acids, or from about 160 amino acids to about 170 amino acids, or from about 180 amino acids to about 190 amino acids, or from about 200 amino acids to about 250 amino acids, or from about 300 amino acids to about 350 amino acids, or from about 400 amino acids to about 450 amino acids, or from about 500 amino acids to about 600 amino acids, to the extent that said length is consistent with the SEQ ID No, or the full length protein, for example any full length protein in SEQ ID NOs: 1-114, may be required for function.

As discussed, the invention includes a novel protein domain, including several examples of THAP-family members. The invention thus encompasses a THAP-family member comprising a polypeptide having at least a THAP domain sequence in the protein or corresponding nucleic acid molecule, preferably a THAP domain sequence

corresponding to SEQ ID NOs: 1-2. A THAP-family member may comprise an amino acid sequence of at least about 25, 30, 35, 40, 45, 50, 60, 70, 80 to 90 amino acid residues in length, of which at least about 50-80%, preferably at least about 60-70%, more preferably at least about 65%, 75% or 90% of the amino acid residues are identical or similar amino acids-to the THAP consensus domain SEQ ID NOs: 1-2.

Identity or similarity may be determined using any desired algorithm, including the algorithms and parameters for determining homology which are described herein.

Optionally, a THAP-domain-containing THAP-family polypeptide comprises a nuclear localization sequence (NLS). As used herein, the term nuclear localization sequence refers to an amino sequence allowing the THAP-family polypeptide to be localized or transported to the cell nucleus. A nuclear localization sequence generally comprises at least about 10, preferably about 13, preferably about 16, more preferably about 19, and even more preferably about 21, 23, 25, 30, 35 or 40 amino acid residues. Alternatively, a THAP-family polypeptide may comprise a deletion of part or the entire NLS or a substitution or insertion in a NLS sequence, such that the modified THAP-family polypeptide is not localized or transported to the cell nucleus.

Isolated proteins of the present invention, preferably THAP family or THAP domain polypeptides, or a biologically active fragments or homologues thereof, have an amino acid sequence sufficiently homologous to the consensus amino acid sequence of SEQ ID NOs: 1-2. As used herein, the term "sufficiently homologous" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share common structural domains or motifs and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains have at least about 30-40% identity, preferably at least about 40-50% identity, more preferably at least about 50-60%, and even more preferably at least about 60-70%, 70-80%, 80%, 90%, 95%, 97%, 98%, 99% or 99.8% identity across the amino acid sequences of the domains and contain at least one and preferably two structural domains or motifs, are defined herein as sufficiently homologous. Furthermore, amino acid or nucleotide sequences which share at least about 30%, preferably at least about 40%, more preferably at least about

60%, 70%, 80%, 90%, 95%, 97%, 98%, 99% or 99.8% identity and share a common functional activity are defined herein as sufficiently homologous.

It be appreciated that the invention encompasses any of the THAP-family polypeptides, as well as fragment thereof, nucleic acids complementary thereto and nucleic acids capable of hybridizing thereto under stringent conditions.

THAP-0 to THAP11

As mentioned, the inventors have identified several THAP-family members, including THAP-0, THAP1, THAP-2, THAP-3, THAP-4, THAP-5, THAP-6, THAP-7, THAP-8, THAP-9, THAP10 and THAP11.

THAP1 Nucleic Acids

The human THAP1 coding sequence, which is approximately 639 nucleotides in length shown in SEQ ID NO: 160, encodes a protein which is approximately 213 amino acid residues in length. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP1 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein.

The human THAP1 gene is localized at chromosomes 8, 18, 11.

The THAP1 protein comprises a THAP domain at amino acids 1-89, the role of which in apoptosis is further demonstrated herein. The THAP1 protein comprises an interferon gamma homology motif at amino acids 136-169 of human THAP1 (NYTVEDTMHQRKRIHQLEQQVEKLRKKLKTAQQR) (SEQ ID NO: 178), exhibiting 41% identity in a 34 residue overlap with human interferon gamma (amino acids 98-131). PML-NBs are closely linked to IFNgamma, and many PML-NB components are induced by IFNgamma, with IFN gamma responsive elements in the promoters of the corresponding genes. The THAP1 protein also includes a nuclear localization sequence at amino acids 146-165 of human THAP1 (RKRIHQLEQQVEKLRKKLKT) (SEQ ID NO: 179). This sequence is responsible for localization of THAP1 in the nucleus. As demonstrated in the examples provided herein, deletion mutants of THAP1 lacking this sequence are no longer localized in the

cell nucleus. The THAP1 protein further comprises a PAR4 binding motif ($LE(X)_{14}$ $QRXRRQXR(X)_{11}QR/KE$) (SEQ ID NO: 180). The core of this motif has been defined experimentally by site directed mutagenesis and by comparison with mouse ZIP/DAP-like kinase (another PAR4 binding partner) it overlaps amino acids 168-175 of human THAP1 but the motif may also include a few residues upstream and downstream.

ESTs corresponding to THAP1 have been identified, and may be specifically included or excluded from the nucleic acids of the invention. The ESTs, as indicated below by accession number, provide evidence for tissue distribution for THAP1 as follows : AL582975 (B cells from Burkitt lymphoma); BG708372 (Hypothalamus); BG563619 (liver); BG497522 (adenocarcinoma); BG616699 (liver); BE932253 (head_neck); AL530396 (neuroblastoma cells).

An object of the invention is a purified, isolated, or recombinant nucleic acid comprising the nucleotide sequence of SEQ ID NO: 160, complementary sequences thereto, and fragments thereof. The invention also pertains to a purified or isolated nucleic acid comprising a polynucleotide having at least 95% nucleotide identity with a polynucleotide of SEQ ID NO: 160, advantageously 99 % nucleotide identity, preferably 99.5% nucleotide identity and most preferably 99.8% nucleotide identity with a polynucleotide of SEQ ID NO: 160, or a sequence complementary thereto or a biologically active fragment thereof. Another object of the invention relates to purified, isolated or recombinant nucleic acids comprising a polynucleotide that hybridizes, under the stringent hybridization conditions defined herein, with a polynucleotide of SEQ ID NO: 160, or a sequence complementary thereto or a variant thereof or a biologically active fragment thereof. In further embodiments, nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID NO: 160, or the complements thereof.

Also encompassed is a purified, isolated, or recombinant nucleic acid polynucleotide encoding a THAP1 polypeptide of the invention, as further described herein.

In another preferred aspect, the invention pertains to purified or isolated nucleic acid molecules that encode a portion or variant of a THAP1 protein, wherein the portion or variant displays a THAP1 activity of the invention. Preferably said portion or variant

is a portion or variant of a naturally occurring full-length THAP1 protein. In one example, the invention provides a polynucleotide comprising, consisting essentially of, or consisting of a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID NO: 160, wherein said nucleic acid encodes a THAP1 portion or variant having a THAP1 activity described herein. In other embodiments, the invention relates to a polynucleotide encoding a THAP1 portion consisting of 8-20, 20-50, 50-70, 60-100, 100 - 150, 150- 200, 200-205 or 205-212 amino acids of SEQ ID NO: 3, or a variant thereof, wherein said THAP1 portion displays a THAP1 activity described herein.

The sequence of SEQ ID NO: 160 corresponds to the human THAP1 cDNA. This cDNA comprises sequences encoding the human THAP1 protein (i.e., "the coding region", from nucleotides 202 to 840, as well as 5' untranslated sequences (nucleotides 1-201) and 3' untranslated sequences (nucleotides 841 to 2173).

Also encompassed by the THAP1 nucleic acids of the invention are nucleic acid molecules which are complementary to THAP1 nucleic acids described herein. Preferably, a complementary nucleic acid is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO: 160, such that it can hybridize to the nucleotide sequence shown in SEQ ID NO: 160, thereby forming a stable duplex.

Another object of the invention is a purified, isolated, or recombinant nucleic acid encoding a THAP1 polypeptide comprising, consisting essentially of, or consisting of the amino acid sequence of SEQ ID NO: 3, or fragments thereof, wherein the isolated nucleic acid molecule encodes one or more motifs selected from the group consisting of a THAP domain, a THAP1 target binding region, a nuclear localization signal and a interferon gamma homology motif. Preferably said THAP1 target binding region is a PAR4 binding region or a DNA binding region. For example, the purified, isolated or recombinant nucleic acid may comprise a genomic DNA or fragment thereof which encodes the polypeptide of SEQ ID NO: 3 or a fragment thereof or a cDNA consisting of, consisting essentially of, or comprising the sequence of SEQ ID NO: 160 or fragments thereof, wherein the isolated nucleic acid molecule encodes one or more motifs selected from the group consisting of a THAP domain, a THAP1-target binding region, a nuclear localization signal and a interferon gamma homology motif. Any combination of said motifs may also be specified. Preferably said THAP1 target

binding region is a PAR4 binding region or a DNA binding region. Particularly preferred nucleic acids of the invention include isolated, purified, or recombinant THAP1 nucleic acids comprising, consisting essentially of, or consisting of a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or 300 nucleotides of a sequence selected from the group consisting of nucleotide positions ranges consisting of 607 to 708, 637 to 696 and 703 to 747 of SEQ ID NO: 160. In preferred embodiments, a THAP1 nucleic acid encodes a THAP1 polypeptide comprising at least two THAP1 functional domains, such as for example a THAP domain and a PAR4 binding region.

In further preferred embodiments, a THAP1 nucleic acid comprises a nucleotide sequence encoding a THAP domain having the consensus amino acid sequence of the formula of SEQ ID NOs: 1-2. A THAP1 nucleic acid may also encode a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids to the THAP domain consensus sequence (SEQ ID NOs: 1-2). The present invention also embodies isolated, purified, and recombinant polynucleotides which encode a polypeptide comprising a contiguous span of at least 6 amino acids, preferably at least 8 or 10 amino acids, more preferably at least 15, 25, 30, 35, 40, 45, 50, 60, 70, 80 or 90 amino acids according to the formula of SEQ ID NO: 1-2.

The nucleotide sequence determined from the cloning of the THAP1 gene allows for the generation of probes and primers designed for use in identifying and/or cloning other THAP1 family members (e.g. sharing the novel functional domains), as well as THAP1 homologues from other species.

A nucleic acid fragment encoding a "biologically active portion of a THAP1 protein" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO: 160, which encodes a polypeptide having a THAP1 biological activity (the biological activities of the THAP1 proteins described herein), expressing the encoded portion of the THAP1 protein (e.g., by recombinant expression in vitro or in vivo) and assessing the activity of the encoded portion of the THAP1 protein.

The invention further encompasses nucleic acid molecules that differ from the THAP1 nucleotide sequences of the invention due to degeneracy of the genetic code and encode the same THAP1 proteins and fragment of the invention.

In addition to the THAP1 nucleotide sequences described above, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the THAP1 proteins may exist within a population (e.g., the human population). Such genetic polymorphism may exist among individuals within a population due to natural allelic variation. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of a THAP1 gene.

Nucleic acid molecules corresponding to natural allelic variants and homologues of the THAP1 nucleic acids of the invention can be isolated based on their homology to the THAP1 nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Probes based on the THAP1 nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g., the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a THAP1 protein, such as by measuring a level of a THAP1-encoding nucleic acid in a sample of cells from a subject e.g., detecting THAP1 mRNA levels or determining whether a genomic THAP1 gene has been mutated or deleted.

THAP1 Polypeptides

The term "THAP1 polypeptides" is used herein to embrace all of the proteins and polypeptides of the present invention. Also forming part of the invention are polypeptides encoded by the polynucleotides of the invention, as well as fusion polypeptides comprising such polypeptides. The invention embodies THAP1 proteins from humans, including isolated or purified THAP1 proteins consisting of, consisting essentially of, or comprising the sequence of SEQ ID NO: 3.

The invention concerns the polypeptide encoded by a nucleotide sequence of SEQ ID NO: 160, a complementary sequence thereof or a fragment thereto.

The present invention embodies isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID NO: 3. In other preferred embodiments the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the THAP1 protein sequence. The invention also concerns the polypeptide encoded by the THAP1 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof.

One aspect of the invention pertains to isolated THAP1 proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-THAP1 antibodies. In one embodiment, native THAP1 proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, THAP1 proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a THAP1 protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

Typically, biologically active portions comprise a domain or motif with at least one activity of the THAP1 protein. The present invention also embodies isolated, purified, and recombinant portions or fragments of one THAP1 polypeptide comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 100 or 200 amino acids of SEQ ID NO: 3. Also encompassed are THAP1 polypeptide which comprise between 10 and 20, between 20 and 50, between 30 and 60, between 50 and 100, or between 100 and 200 amino acids of SEQ ID NO: 3. In other preferred embodiments the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the THAP1 protein sequence.

A biologically active THAP1 protein may, for example, comprise at least 1, 2, 3, 5, 10, 20 or 30 amino acid changes from the sequence of SEQ ID NO: 3, or may encode a biologically active THAP1 protein comprising at least 1%, 2%, 3%, 5%, 8%, 10% or 15% changes in amino acids from the sequence of SEQ ID NO: 3.

In a preferred embodiment, the THAP1 protein comprises, consists essentially of, or consists of a THAP domain at amino acid positions 1 to 89 shown in SEQ ID NO: 3, or fragments or variants thereof. In other aspects, a THAP1 polypeptide comprises a THAP1-target binding region, a nuclear localization signal and/or a Interferon Gamma Homology Motif. Preferably a THAP1 target binding region is a PAR4 binding region or a DNA binding region. The invention also concerns the polypeptide encoded by the THAP1 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80, 90 or 100 amino acids of an amino acid sequence selected from the group consisting of positions 1 to 90, 136 to 169, 146 to 165 and 168 to 175 of SEQ ID NO: 3. In another aspect, a THAP1 polypeptide may encode a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids to the THAP domain consensus sequence (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP1 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 90 shown in SEQ ID NO: 3, or fragments or variants thereof.

In other embodiments, the THAP1 protein is substantially homologous to the sequences of SEQ ID NO: 3, and retains the functional activity of the THAP1 protein, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described further herein. Accordingly, in another embodiment, the THAP1 protein is a protein which comprises an amino acid sequence shares more than about 60% but less than 100% homology with the amino acid sequence of SEQ ID NO: 3 and retains the functional activity of the THAP1 proteins of SEQ ID NO: 3, respectively. Preferably, the protein is at least about 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 92%, 95%, 97%, 98%, 99% or 99.8% homologous to SEQ ID NO: 3, but is not identical to SEQ ID NO: 3. Preferably the THAP1 is less than identical (e.g. 100% identity) to a naturally occurring THAP1. Percent homology can be determined as further detailed above.

THAP-2 to THAP11 and THAP-0 Nucleic Acids

As mentioned, the invention provides several members of the THAP-family. THAP-2, THAP-3, THAP-4, THAP-5, THAP-6, THAP-7, THAP-8, THAP-9, THAP10, THAP11 and THAP-0 are described herein. The human and mouse nucleotide sequences corresponding to the human cDNA sequences are listed in SEQ ID NOs: 161-171; and the human amino acid sequences are listed respectively in SEQ ID NOs: 4-14. Also encompassed by the invention are orthologs of said THAP-family sequences, including mouse, rat, pig and other orthologs, the amino acid sequences of which are listed in SEQ ID NOs: 16-114 and the cDNA sequences are listed in SEQ ID NOs: 172-175.

THAP-2

The human THAP-2 cDNA, which is approximately 1302 nucleotides in length shown in SEQ ID NO: 161, encodes a protein which is approximately 228 amino acid residues in length, shown in SEQ ID NO: 4. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP-2 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The human THAP-2 gene is localized at chromosomes 12 and 3. The THAP-2 protein comprises a THAP domain at amino acids 1 to 89. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP-2 is expressed as follows: BG677995 (squamous cell carcinoma); AV718199 (hypothalamus); BI600215 (hypothalamus); AI208780 (Soares_testis_NHT); BE566995 (carcinoma cell line); AI660418 (thymus pooled)

THAP-3

The human THAP-3 cDNA which is approximately 1995 nucleotides in length shown in SEQ ID NO: 162. The THAP-3 gene encodes a protein which is approximately 239 amino acid residues in length, shown in SEQ ID NO: 5. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP-3 proteins or biologically active portions thereof as further described herein, as

well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The human THAP-3 gene is localized at chromosome 1. The THAP-3 protein comprises a THAP domain at amino acids 1 to 89. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP-3 is expressed as follows: BG700517 (hippocampus); BI460812 (testis); BG707197 (hypothalamus); AW960428 (-); BG437177 (large cell carcinoma); BE962820 (adenocarcinoma); BE548411 (cervical carcinoma cell line); AL522189 (neuroblastoma cells); BE545497 (cervical carcinoma cell line); BE280538 (choriocarcinoma); BI086954 (cervix); BE744363 (adenocarcinoma cell line); and BI549151 (hippocampus).

THAP-4

The human THAP-4cDNA, shown as a sequence having 1999 nucleotides in length shown in SEQ ID NO: 163, encodes a protein which is approximately 577 amino acid residues in length, shown in SEQ ID NO: 6. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP-4 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The THAP-4 protein comprises a THAP domain at amino acids 1 to 90. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP-4 is expressed as follows: AL544881 (placenta); BE384014 (melanotic melanoma); AL517205 (neuroblastoma cells); BG394703 (retinoblastoma); BG472327 (retinoblastoma); BI196071 (neuroblastoma); BE255202 (retinoblastoma); BI017349 (lung_tumor); BF972153 (leiomyosarcoma cell line); BG116061 (duodenal adenocarcinoma cell line); AL530558 (neuroblastoma cells); AL520036 (neuroblastoma cells); AL559902 (B cells from Burkitt lymphoma); AL534539 (Fetal brain); BF686560 (leiomyosarcoma cell line); BF345413 (anaplastic oligodendroglioma with 1p/19q loss); BG117228 (adenocarcinoma cell line); BG490646 (large cell carcinoma); and BF769104 (epid_tumor).

THAP-5

The human THAP-5 cDNA, shown as a sequence having 1034 nucleotides in length shown in SEQ ID NO: 164, encodes a protein which is approximately 239 amino acid residues in length, shown in SEQ ID NO: 7. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP-5 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The human THAP-5 gene is localized at chromosome 7. The THAP-5 protein comprises a THAP domain at amino acids 1 to 90. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP-5 is expressed as follows: BG575430 (mammary adenocarcinoma cell line); BI545812 (hippocampus); BI560073 (testis); BG530461 (embryonal carcinoma); BF244164 (glioblastoma); BI461364 (testis); AW407519 (germinal center B cells); BF103690 (embryonal carcinoma); and BF939577 (kidney).

THAP-6

The human THAP-6cDNA, shown as a sequence having 2291 nucleotides in length shown in SEQ ID NO: 165, encodes a protein which is approximately 222 amino acid residues in length, shown in SEQ ID NO: 8. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP-6 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The human THAP-6 gene is localized at chromosome 4. The THAP-6 protein comprises a THAP domain at amino acids 1 to 90. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP-6 is expressed as follows: AV684783 (hepatocellular carcinoma); AV698391 (hepatocellular carcinoma); BI560555 (testis); AV688768 (hepatocellular carcinoma); AV692405 (hepatocellular carcinoma); and AV696360 (hepatocellular carcinoma).

THAP-7

The human THAP-7 cDNA, shown as a sequence having 1242 nucleotides in length shown in SEQ ID NO: 166, encodes a protein which is approximately 309 amino acid residues in length, shown in SEQ ID NO: 9. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP-7 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The human THAP-7 gene is localized at chromosome 22q11.2. The THAP-7 protein comprises a THAP domain at amino acids 1 to 90. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP-7 is expressed as follows: BI193682 (epithelioid carcinoma cell line); BE253146 (retinoblastoma); BE622113 (melanotic melanoma); BE740360 (adenocarcinoma cell line); BE513955 (Burkitt lymphoma); AL049117 (testis); BF952983 (nervous_normal); AW975614 (-); BE273270 (renal cell adenocarcinoma); BE738428 (glioblastoma); BE388215 (endometrium adenocarcinoma cell line); BF762401 (colon_est); and BG329264 (retinoblastoma).

THAP-8

The human THAP-8 cDNA, shown as a sequence having 1383 nucleotides in length shown in SEQ ID NO: 167, encodes a protein which is approximately 274 amino acid residues in length, shown in SEQ ID NO: 10. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP-8 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The human THAP-8 gene is localized at chromosome 19. The THAP-8 protein comprises a THAP domain at amino acids 1 to 92. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP-8 is expressed as follows: BG703645 (hippocampus);

BF026346 (melanotic melanoma); BE728495 (melanotic melanoma); BG334298 (melanotic melanoma); and BE390697 (endometrium adenocarcinoma cell line).

THAP-9

The human THAP-9 cDNA, shown as a sequence having 693 nucleotides in length shown in SEQ ID NO: 168, encodes a protein which is approximately 231 amino acid residues in length, shown in SEQ ID NO: 11. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP-9 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The THAP-9 protein comprises a THAP domain at amino acids 1 to 92. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP-9 is expressed as follows: AA333595 (Embryo 8 weeks).

THAP10

The human THAP10 cDNA, shown as a sequence having 771 nucleotides in length shown in SEQ ID NO: 169, encodes a protein which is approximately 257 amino acid residues in length, shown in SEQ ID NO: 12. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP10 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The human THAP10 gene is localized at chromosome 15. The THAP10 protein comprises a THAP domain at amino acids 1 to 90. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP10 is expressed as follows: AL526710 (neuroblastoma cells); AV725499 (Hypothalamus); AW966404 (-); AW296810 (lung); and AL557817 (T cells from T cell leukemia).

THAP11

The human THAP11 cDNA, shown as a sequence having 942 nucleotides in length shown in SEQ ID NO: 170, encodes a protein which is approximately 314 amino acid residues in length, shown in SEQ ID NO: 13. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP11 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The human THAP11 gene is localized at chromosome 16. The THAP11 protein comprises a THAP domain at amino acids 1 to 90. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP11 is expressed as follows: AU142300 (retinoblastoma); BI261822 (lymphoma cell line); BG423102 (renal cell adenocarcinoma); and BG423864 (kidney).

THAP-0

The human THAP-0 cDNA, shown as a sequence having 2283 nucleotides in length shown in SEQ ID NO: 171, encodes a protein which is approximately 761 amino acid residues in length, shown in SEQ ID NO: 14. One aspect of the invention pertains to purified or isolated nucleic acid molecules that encode THAP-0 proteins or biologically active portions thereof as further described herein, as well as nucleic acid fragments thereof. Said nucleic acids may be used for example in therapeutic methods and drug screening assays as further described herein. The human THAP-0 gene is localized at chromosome 11. The THAP-0 protein comprises a THAP domain at amino acids 1 to 90. Analysis of expressed sequences (accession numbers indicated, which may be specifically included or excluded from the nucleic acids of the invention) in databases suggests that THAP-0 is expressed as follows: BE713222 (head_neck); BE161184 (head_neck); AL119452 (amygdala); AU129709 (teratocarcinoma); AW965460 (-); AW965460(-); AW958065 (-); and BE886885 (leiomyosarcoma).

An object of the invention is a purified, isolated, or recombinant nucleic acid comprising the nucleotide sequence of SEQ ID NOs: 161-171, 173-175 or complementary sequences thereto, and fragments thereof. The invention also pertains to

a purified or isolated nucleic acid comprising a polynucleotide having at least 95% nucleotide identity with a polynucleotide of SEQ ID NOs: 161-171 or 173-175, advantageously 99 % nucleotide identity, preferably 99.5% nucleotide identity and most preferably 99.8% nucleotide identity with a polynucleotide of SEQ ID NOs: 161-171, 173-175 or a sequence complementary thereto or a biologically active fragment thereof. Another object of the invention relates to purified, isolated or recombinant nucleic acids comprising a polynucleotide that hybridizes, under the stringent hybridization conditions defined herein, with a polynucleotide of SEQ ID NOs: 161-171, 173-175 or a sequence complementary thereto or a variant thereof or a biologically active fragment thereof. In further embodiments, nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of a sequence selected from the group consisting of SEQ ID NOs: 161-171, 173-175 or the complements thereof.

Also encompassed is a purified, isolated, or recombinant nucleic acid polynucleotide encoding a THAP-2 to THAP11 or THAP-0 polypeptide of the invention, as further described herein.

In another preferred aspect, the invention pertains to purified or isolated nucleic acid molecules that encode a portion or variant of a THAP-2 to THAP11 or THAP-0 protein, wherein the portion or variant displays a THAP-2 to THAP11 or THAP-0 activity of the invention. Preferably said portion or variant is a portion or variant of a naturally occurring full-length THAP-2 to THAP11 or THAP-0 protein. In one example, the invention provides a polynucleotide comprising, consisting essentially of, or consisting of a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides, to the extent that the length of said span is consistent with the length of the SEQ ID NO, of a sequence selected from the group consisting of SEQ ID NOs: 161-171, 173-175, wherein said nucleic acid encodes a THAP-2 to THAP11 or THAP-0 portion or variant having a THAP-2 to THAP11 or THAP-0 activity described herein. In other embodiment, the invention relates to a polynucleotide encoding a THAP-2 to THAP11 or THAP-0 portion consisting of 8-20, 20-50, 50-70, 60-100, 100 - 150, 150- 200, 200-250 or 250 - 350 amino acids, to the extent that the length of said portion is consistent with the length of the SEQ ID NO: of

a sequence selected from the group consisting of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98, 100-114 or a variant thereof, wherein said THAP-2 to THAP11 or THAP-0 portion displays a THAP-2 to THAP11 or THAP-0 activity described herein.

A THAP-2 to THAP11 or THAP-0 variant nucleic acid may, for example, encode a biologically active THAP-2 to THAP11 or THAP-0 protein comprising at least 1, 2, 3, 5, 10, 20 or 30 amino acid changes from the respective sequence selected from the group consisting of SEQ ID NO: 4-14, 17-21, 23-40, 42-56, 58-98 and 100-114 or may encode a biologically active THAP-2 to THAP11 or THAP-0 protein comprising at least 1%, 2%, 3%, 5%, 8%, 10% or 15% changes in amino acids from the respective sequence of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 and 100-114.

The sequences of SEQ ID NOs: 4-14 correspond to the human THAP-2 to THAP11 and THAP-0 DNAs respectively. SEQ ID NOs: 17-21, 23-40, 42-56, 58-98, 100-114 correspond to mouse, rat, pig and other orthologs.

Also encompassed by the THAP-2 to THAP11 and THAP-0 nucleic acids of the invention are nucleic acid molecules which are complementary to THAP-2 to THAP11 or THAP-0 nucleic acids described herein. Preferably, a complementary nucleic acid is sufficiently complementary to the nucleotide respective sequence shown in SEQ ID NOs: 161-171 and 173-175 such that it can hybridize to said nucleotide sequence shown in SEQ ID NOs: 161-171 and 173-175, thereby forming a stable duplex.

Another object of the invention is a purified, isolated, or recombinant nucleic acid encoding a THAP-2 to THAP11 or THAP-0 polypeptide comprising, consisting essentially of, or consisting of an amino acid sequence selected from the group consisting of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98, 100-114 or fragments thereof, wherein the isolated nucleic acid molecule encodes a THAP domain or a THAP-2 to THAP11 or THAP-0 target binding region. Preferably said target binding region is a protein binding region, preferably a PAR-4 binding region, or preferably said target binding region is a DNA binding region. For example, the purified, isolated or recombinant nucleic acid may comprise a genomic DNA or fragment thereof which encodes a polypeptide having a sequence selected from the group consisting of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98, 100-114 or a fragment thereof. The purified, isolated or recombinant nucleic acid may alternatively comprise a cDNA consisting of, consisting essentially of, or comprising a sequence selected from the group consisting

of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98, 100-114 or fragments thereof, wherein the isolated nucleic acid molecule encodes a THAP domain or a THAP-2 to THAP11 or THAP-0 target binding region. In preferred embodiments, a THAP-2 to THAP11 or THAP-0 nucleic acid encodes a THAP-2 to THAP11 or THAP-0 polypeptide comprising at least two THAP-2 to THAP11 or THAP-0 functional domains, such as for example a THAP domain and a THAP-2 to THAP11 or THAP-0 target binding region.

Particularly preferred nucleic acids of the invention include isolated, purified, or recombinant THAP-2 to THAP11 or THAP-0 nucleic acids comprising, consisting essentially of, or consisting of a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200 or 250 nucleotides of a sequence selected from the group consisting of nucleotide positions coding for the relevant amino acids as given in the SEQ ID NO: 161-171 and 173-175.

In further preferred embodiments, a THAP-2 to THAP11 or THAP-0 nucleic acid comprises a nucleotide sequence encoding a THAP domain having the consensus amino acid sequence of the formula of SEQ ID NOs: 1-2. A THAP-2 to THAP11 or THAP-0 nucleic acid may also encode a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids-to the THAP consensus domain (SEQ ID NOs: 1-2). The present invention also embodies isolated, purified, and recombinant polynucleotides which encode a polypeptide comprising a contiguous span of at least 6 amino acids, preferably at least 8 or 10 amino acids, more preferably at least 15, 25, 30, 35, 40, 45, 50, 60, 70, 80 or 90 amino acids of SEQ ID NOs: 1-2 .

The nucleotide sequence determined from the cloning of the THAP-2 to THAP11 or THAP-0 genes allows for the generation of probes and primers designed for use in identifying and/or cloning other THAP family members, particularly sequences related to THAP-2 to THAP11 or THAP-0 (e.g. sharing the novel functional domains), as well as THAP-2 to THAP11 or THAP-0 homologues from other species.

A nucleic acid fragment encoding a biologically active portion of a THAP-2 to THAP11 or THAP-0 protein can be prepared by isolating a portion of a nucleotide sequence selected from the group consisting of SEQ ID NOs: 161-171 and 173-175, which encodes a polypeptide having a THAP-2 to THAP11 or THAP-0 biological

activity (the biological activities of the THAP-family proteins described herein), expressing the encoded portion of the THAP-2 to THAP11 or THAP-0 protein (e.g., by recombinant expression in vitro or in vivo) and assessing the activity of the encoded portion of the THAP-2 to THAP11 or THAP-0 protein.

The invention further encompasses nucleic acid molecules that differ from the THAP-2 to THAP11 or THAP-0 nucleotide sequences of the invention due to degeneracy of the genetic code and encode the same THAP-2 to THAP11 or THAP-0 protein, or fragment thereof, of the invention.

In addition to the THAP-2 to THAP11 or THAP-0 nucleotide sequences described above, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the respective THAP-2 to THAP11 or THAP-0 protein may exist within a population (e.g., the human population). Such genetic polymorphism may exist among individuals within a population due to natural allelic variation. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of a particular THAP-2 to THAP11 or THAP-0 gene.

Nucleic acid molecules corresponding to natural allelic variants and homologues of the THAP-2 to THAP11 or THAP-0 nucleic acids of the invention can be isolated based on their homology to the THAP-2 to THAP11 or THAP-0 nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Probes based on the THAP-2 to THAP11 or THAP-0 nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g., the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a THAP-2 to THAP11 or THAP-0 protein, such as by measuring a level of a THAP-2 to THAP11 or THAP-0-encoding nucleic acid in a sample of cells from a subject e.g., detecting THAP-2 to THAP11 or THAP-0 mRNA levels or determining whether a genomic THAP-2 to THAP11 or THAP-0 gene has been mutated or deleted.

THAP-2 to THAP11 and THAP-0 Polypeptides

The term "THAP-2 to THAP11 or THAP-0 polypeptides" is used herein to embrace all of the proteins and polypeptides of the present invention relating to THAP-2, THAP-3, THAP-4, THAP-5, THAP-6, THAP-7, THAP-8, THAP-9, THAP10, THAP11 and THAP-0. Also forming part of the invention are polypeptides encoded by the polynucleotides of the invention, as well as fusion polypeptides comprising such polypeptides. The invention embodies THAP-2 to THAP11 or THAP-0 proteins from humans, including isolated or purified THAP-2 to THAP11 or THAP-0 proteins consisting of, consisting essentially of, or comprising a sequence selected from the group consisting of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 and 100-114.

The invention concerns the polypeptide encoded by a nucleotide sequence selected from the group consisting of SEQ ID NOs: 161-171, 172-175 and a complementary sequence thereof and a fragment thereof.

The present invention embodies isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 100, 150, 200, 300 or 500 amino acids, to the extent that said span is consistent with the particular SEQ ID NO:, of a sequence selected from the group consisting of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 and 100-114. In other preferred embodiments the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the THAP-2 to THAP11 or THAP-0 protein sequence.

One aspect of the invention pertains to isolated THAP-2 to THAP11 and THAP-0 proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-THAP-2 to THAP11 or THAP-0 antibodies. In one embodiment, native THAP-2 to THAP11 or THAP-0 proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, THAP-2 to THAP11 or THAP-0 proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a THAP-2 to THAP11 or THAP-0 protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

Biologically active portions of a THAP-2 to THAP11 or THAP-0 protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the THAP-2 to THAP11 or THAP-0 protein, e.g., an amino acid sequence shown in SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 or 100-114, which include less amino acids than the respective full length THAP-2 to THAP11 or THAP-0 protein, and exhibit at least one activity of the THAP-2 to THAP11 or THAP-0 protein. The present invention also embodies isolated, purified, and recombinant portions or fragments of a THAP-2 to THAP11 or THAP-0 polypeptide comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 100,150, 200, 300 or 500 amino acids, to the extent that said span is consistent with the particular SEQ ID NO, of a sequence selected from the group consisting of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 and 100-114. Also encompassed are THAP-2 to THAP11 or THAP-0 polypeptides which comprise between 10 and 20, between 20 and 50, between 30 and 60, between 50 and 100, or between 100 and 200 amino acids of a sequence selected from the group consisting of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 and 100-114. In other preferred embodiments the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the THAP-2 to THAP11 or THAP-0 protein sequence.

A biologically active THAP-2 to THAP11 or THAP-0 protein may, for example, comprise at least 1, 2, 3, 5, 10, 20 or 30 amino acid changes from the sequence of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 or 100-114, or may encode a biologically active THAP-2 to THAP11 or THAP-0 protein comprising at least 1%, 2%, 3%, 5%, 8%, 10% or 15% changes in amino acids from the sequence of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 or 100-114.

In a preferred embodiment, the THAP-2 protein comprises, consists essentially of, or consists of a THAP-2 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 89 shown in SEQ ID NO: 4, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP-2 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a

contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 89 amino acids of a sequence comprising amino acid positions 1 to 89 of SEQ ID NO: 4. In another aspect, a THAP-2 polypeptide may comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids-to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP-2 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 89 shown in SEQ ID NO: 4, or fragments or variants thereof. Preferably, said THAP-2 polypeptide comprises a PAR-4 binding domain and/or a DNA binding domain.

In a preferred embodiment, the THAP-3 protein comprises, consists essentially of, or consists of a THAP-3 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 89 shown in SEQ ID NO: 5, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP-3 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 89 amino acids of a sequence comprising amino acid positions 1 to 89 of SEQ ID NO: 5. In another aspect, a THAP-3 polypeptide may comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids-to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP-3 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 89 shown in SEQ ID NO: 5, or fragments or variants thereof. Preferably, said THAP-3 polypeptide comprises a PAR-4 binding domain and/or a DNA binding domain.

In a preferred embodiment, the THAP-4 protein comprises, consists essentially of, or consists of a THAP-4 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 90 shown in SEQ ID NO: 6, or fragments or variants

thereof. The invention also concerns the polypeptide encoded by the THAP-4 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 90 amino acids of a sequence comprising amino acid positions 1 to 90 of SEQ ID NO: 6. In another aspect, a THAP-4 polypeptide may comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids-to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP-4 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 90 shown in SEQ ID NO: 6, or fragments or variants thereof.

In a preferred embodiment, the THAP-5 protein comprises, consists essentially of, or consists of a THAP-5 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 90 shown in SEQ ID NO: 7, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP-5 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 90 amino acids of a sequence comprising amino acid positions 1 to 90 of SEQ ID NO: 7. In another aspect, a THAP-5 polypeptide may comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids-to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP-5 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 90 shown in SEQ ID NO: 7, or fragments or variants thereof.

In a preferred embodiment, the THAP-6 protein comprises, consists essentially of, or consists of a THAP-6 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 90 shown in SEQ ID NO: 8, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP-6 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 90 amino acids of a sequence comprising amino acid positions 1 to 90 of SEQ ID NO: 8. In another aspect, a THAP-6 polypeptide may comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids-to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP-6 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 90 shown in SEQ ID NO: 8, or fragments or variants thereof.

In a preferred embodiment, the THAP-7 protein comprises, consists essentially of, or consists of a THAP-7 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 90 shown in SEQ ID NO: 9, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP-7 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 90 amino acids of a sequence comprising amino acid positions 1 to 90 of SEQ ID NO: 9. In another aspect, a THAP-7 polypeptide may comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids-to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP-7 polypeptide comprising, consisting

essentially of, or consisting of a THAP domain at amino acid positions 1 to 90 shown in SEQ ID NO: 9, or fragments or variants thereof.

In a preferred embodiment, the THAP-8 protein comprises, consists essentially of, or consists of a THAP-8 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 92 shown in SEQ ID NO: 10, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP-8 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 90 amino acids of a sequence comprising amino acid positions 1 to 92 of SEQ ID NO: 10. In another aspect, a THAP-8 polypeptide may comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids–to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP-8 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 92 shown in SEQ ID NO: 10, or fragments or variants thereof.

In a preferred embodiment, the THAP-9 protein comprises, consists essentially of, or consists of a THAP-9 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 92 shown in SEQ ID NO: 11, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP-9 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 90 amino acids of a sequence comprising amino acid positions 1 to 92 of SEQ ID NO: 11. In another aspect, a THAP-9 polypeptide may comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids–to the THAP domain

consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP-9 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 92 shown in SEQ ID NO: 11, or fragments or variants thereof.

In a preferred embodiment, the THAP10 protein comprises, consists essentially of, or consists of a THAP10 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 90 shown in SEQ ID NO: 12, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP10 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 90 amino acids of a sequence comprising amino acid positions 1 to 90 of SEQ ID NO: 12. In another aspect, a THAP10 polypeptide may comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids–to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP10 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 90 shown in SEQ ID NO: 12, or fragments or variants thereof.

In a preferred embodiment, the THAP11 protein comprises, consists essentially of, or consists of a THAP11 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 90 shown in SEQ ID NO: 13, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP11 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 90 amino acids of a sequence comprising amino acid positions 1 to 90 of SEQ ID NO: 13. In another aspect, a THAP11 polypeptide may comprise a THAP domain wherein at least about 95%, 90%,

85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids–to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP11 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 90 shown in SEQ ID NO: 13, or fragments or variants thereof.

In a preferred embodiment, the THAP-0 protein comprises, consists essentially of, or consists of a THAP-0 THAP domain, preferably having the amino acid sequence of amino acid positions 1 to 90 shown in SEQ ID NO: 14, or fragments or variants thereof. The invention also concerns the polypeptide encoded by the THAP-0 nucleotide sequences of the invention, or a complementary sequence thereof or a fragment thereof. The present invention thus also embodies isolated, purified, and recombinant polypeptides comprising, consisting essentially of or consisting of a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 70, 80 or 90 amino acids of a sequence comprising amino acid positions 1 to 90 of SEQ ID NO: 14. In another aspect, a THAP-0 polypeptide may comprise a THAP domain wherein at least about 95%, 90%, 85%, 50-80%, preferably at least about 60-70%, more preferably at least about 65% of the amino acid residues are identical or similar amino acids–to the THAP domain consensus domain (SEQ ID NOs: 1-2). Also encompassed by the present invention are isolated, purified, nucleic acids encoding a THAP-0 polypeptide comprising, consisting essentially of, or consisting of a THAP domain at amino acid positions 1 to 90 shown in SEQ ID NO: 14, or fragments or variants thereof.

In other embodiments, the THAP-2 to THAP11 or THAP-0 protein is substantially homologous to the sequences of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 or 100-114 and retains the functional activity of the THAP-2 to THAP11 or THAP-0 protein, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described further herein. Accordingly, in another embodiment, the THAP-2 to THAP11 or THAP-0 protein is a protein which comprises an amino acid sequence that shares more than about 60% but less than 100% homology with the amino acid sequence of SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 or 100-114 and retains the functional activity of the THAP-2 to THAP11 or THAP-0 proteins of SEQ

ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 or 100-114, respectively. Preferably, the protein is at least about 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 92%, 95%, 97%, 98%, 99% or 99.8% homologous to SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 or 100-114, but is not identical to SEQ ID NOs: 4-14, 17-21, 23-40, 42-56, 58-98 or 100-114. Preferably the THAP-2 to THAP11 or THAP-0 is less than identical (e.g. 100% identity) to a naturally occurring THAP-2 to THAP11 or THAP-0. Percent homology can be determined as further detailed above.

Assessing polypeptides, methods for obtaining variant nucleic acids and polypeptides

It will be appreciated that by characterizing the function of THAP-family polypeptides, the invention further provides methods of testing the activity of, or obtaining, functional fragments and variants of THAP-family and THAP domain nucleotide sequences involving providing a variant or modified THAP-family or THAP domain nucleic acid and assessing whether a polypeptide encoded thereby displays a THAP-family activity of the invention. Encompassed is thus a method of assessing the function of a THAP-family or THAP domain polypeptide comprising : (a) providing a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof; and (b) testing said THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof for a THAP-family activity. Any suitable format may be used, including cell free, cell-based and in vivo formats. For example, said assay may comprise expressing a THAP-family or THAP domain nucleic acid in a host cell, and observing THAP-family activity in said cell. In another example, a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof is introduced to a cell, and a THAP-family activity is observed. THAP-family activity may be any activity as described herein, including: (1) mediating apoptosis or cell proliferation when expressed or introduced into a cell, most preferably inducing or enhancing apoptosis, and/or most preferably reducing cell proliferation; (2) mediating apoptosis or cell proliferation of an endothelial cell; (3) mediating apoptosis or cell proliferation of a hyperproliferative cell; (4) mediating apoptosis or cell proliferation of a CNS cell, preferably a neuronal or glial cell; or (5) an activity determined in an animal selected from the group consisting of mediating,

preferably inhibiting angiogenesis, mediating, preferably inhibiting inflammation, inhibition of metastatic potential of cancerous tissue, reduction of tumor burden, increase in sensitivity to chemotherapy or radiotherapy, killing a cancer cell, inhibition of the growth of a cancer cell, or induction of tumor regression.

In addition to naturally-occurring allelic variants of the THAP-family or THAP domain sequences that may exist in the population, the skilled artisan will appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NOs: 160-171, thereby leading to changes in the amino acid sequence of the encoded THAP-family or THAP domain proteins, with or without altering the functional ability of the THAP-family or THAP domain proteins.

Several types of variants are contemplated including 1) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue and such substituted amino acid residue may or may not be one encoded by the genetic code, or 2) one in which one or more of the amino acid residues includes a substituent group, or 3) one in which the mutated THAP-family or THAP domain polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or 4) one in which the additional amino acids are fused to the mutated THAP-family or THAP domain polypeptide, such as a leader or secretory sequence or a sequence which is employed for purification of the mutated THAP-family or THAP domain polypeptide or a preprotein sequence. Such variants are deemed to be within the scope of those skilled in the art.

For example, nucleotide substitutions leading to amino acid substitutions can be made in the sequences of SEQ ID NOs: 160-175 that do not substantially change the biological activity of the protein. An amino acid residue can be altered from the wild-type sequence encoding a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof without altering the biological activity. In general, amino acid residues that are conserved among the THAP-family of THAP domain-containing proteins of the present invention, are predicted to be less amenable to alteration. Furthermore, additional conserved amino acid residues may be amino acids that are conserved between the THAP-family proteins of the present invention.

In one aspect, the invention pertains to nucleic acid molecules encoding THAP family or THAP domain polypeptides, or biologically active fragments or homologues

thereof that contain changes in amino acid residues that are not essential for activity. Such THAP-family proteins differ in amino acid sequence from SEQ ID NOs: 1-114 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 60% homologous to an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114. Preferably, the protein encoded by the nucleic acid molecule is at least about 65-70% homologous to an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114, more preferably sharing at least about 75-80% identity with an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114, even more preferably sharing at least about 85%, 90%, 92%, 95%, 97%, 98%, 99% or 99.8% identity with an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114.

In another aspect, the invention pertains to nucleic acid molecules encoding THAP-family proteins that contain changes in amino acid residues that result in increased biological activity, or a modified biological activity. In another aspect, the invention pertains to nucleic acid molecules encoding THAP-family proteins that contain changes in amino acid residues that are essential for a THAP-family activity. Such THAP-family proteins differ in amino acid sequence from SEQ ID NOs: 1-114 and display reduced or essentially lack one or more THAP-family biological activities. The invention also encompasses a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof which may be useful as dominant negative mutant of a THAP family or THAP domain polypeptide.

An isolated nucleic acid molecule encoding a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof homologous to a protein of any one of SEQ ID NOs: 1-114 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NOs: 1-114 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into any of SEQ ID NOs: 1-114, by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. For example, conservative amino acid substitutions may be made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an

amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof may be replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a THAP-family or THAP domain coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for THAP-family biological activity to identify mutants that retain activity. Following mutagenesis of one of SEQ ID NOs: 1-114, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof encoded by a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof of THAP domain nucleic acid of the invention can be assayed for a THAP-family activity in any suitable assay, examples of which are provided herein.

The invention also provides THAP-family or THAP domain chimeric or fusion proteins. As used herein, a THAP-family or THAP domain "chimeric protein" or "fusion protein" comprises a THAP-family or THAP domain polypeptide of the invention operatively linked, preferably fused in frame, to a non-THAP-family or non-THAP domain polypeptide. In a preferred embodiment, a THAP-family or THAP domain fusion protein comprises at least one biologically active portion of a THAP-family or THAP domain protein. In another preferred embodiment, a THAP-family fusion protein comprises at least two biologically active portions of a THAP-family protein. For example, in one embodiment, the fusion protein is a GST-THAP-family fusion protein in which the THAP-family sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant

THAP-family polypeptides. In another embodiment, the fusion protein is a THAP-family protein containing a heterologous signal sequence at its N-terminus, such as for example to allow for a desired cellular localization in a certain host cell.

The THAP-family or THAP domain fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject in vivo. Moreover, the THAP-family-fusion or THAP domain proteins of the invention can be used as immunogens to produce anti-THAP-family or anti or THAP domain antibodies in a subject, to purify THAP-family or THAP domain ligands and in screening assays to identify molecules which inhibit the interaction of THAP-family or THAP domain with a THAP-family or THAP domain target molecule.

Furthermore, isolated peptidyl portions of the subject THAP-family or THAP domain proteins can also be obtained by screening peptides recombinantly produced from the corresponding fragment of the nucleic acid encoding such peptides. In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, a THAP-family or THAP domain protein of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or preferably divided into overlapping fragments of a desired length. The fragments can be produced (recombinantly or by chemical synthesis) and tested to identify those peptidyl fragments which can function as either agonists or antagonists of a THAP-family protein activity, such as by microinjection assays or in vitro protein binding assays. In an illustrative embodiment, peptidyl portions of a THAP-family protein, such as a THAP domain or a THAP-family target binding region (e.g. PAR4 in the case of THAP1, THAP-2 and THAP-3), can be tested for THAP-family activity by expression as thioredoxin fusion proteins, each of which contains a discrete fragment of the THAP-family protein (see, for example, U.S. Patents 5,270,181 and 5,292,646; and PCT publication W094/02502, the disclosures of which are incorporated herein by reference).

The present invention also pertains to variants of the THAP-family or THAP domain proteins which function as either THAP-family or THAP domain mimetics or as THAP-family or THAP domain inhibitors. Variants of the THAP-family or THAP domain proteins can be generated by mutagenesis, e.g., discrete point mutation or truncation of a THAP-family or THAP domain protein. An agonist of a THAP-family

or THAP domain protein can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a THAP-family or THAP domain protein. An antagonist of a THAP-family or THAP domain protein can inhibit one or more of the activities of the naturally occurring form of the THAP-family or THAP domain protein by, for example, competitively inhibiting the association of a THAP-family or THAP domain protein with a THAP-family target molecule. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, variants of a THAP-family or THAP domain protein which function as either THAP-family or THAP domain agonists (mimetics) or as THAP-family or THAP domain antagonists can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of a THAP-family or THAP domain protein for THAP-family or THAP domain protein agonist or antagonist activity. In one embodiment, a variegated library of THAP-family variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of THAP-family variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential THAP-family sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of THAP-family sequences therein. There are a variety of methods which can be used to produce libraries of potential THAP-family variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential THAP-family sequences.

In addition, libraries of fragments of a THAP-family or THAP domain protein coding sequence can be used to generate a variegated population of THAP-family or THAP domain fragments for screening and subsequent selection of variants of a THAP-family or THAP domain protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a THAP-family coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to

form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the THAP-family protein.

Modified THAP-family or THAP domain proteins can be used for such purposes as enhancing therapeutic or prophylactic efficacy, or stability (e.g., ex vivo shelf life and resistance to proteolytic degradation in vivo). Such modified peptides, when designed to retain at least one activity of the naturally occurring form of the protein, are considered functional equivalents of the THAP-family or THAP domain protein described in more detail herein. Such modified peptide can be produced, for instance, by amino acid substitution, deletion, or addition.

Whether a change in the amino acid sequence of a peptide results in a functional THAP-family or THAP domain homolog (e.g. functional in the sense that it acts to mimic or antagonize the wild-type form) can be readily determined by assessing the ability of the variant peptide to produce a response in cells in a fashion similar to the wild-type THAP-family or THAP domain protein or competitively inhibit such a response. Peptides in which more than one replacement has taken place can readily be tested in the same manner.

This invention further contemplates a method of generating sets of combinatorial mutants of the presently disclosed THAP-family or THAP domain proteins, as well as truncation and fragmentation mutants, and is especially useful for identifying potential variant sequences which are functional in binding to a THAP-family- or THAP domain-target protein but differ from a wild-type form of the protein by, for example, efficacy, potency and/or intracellular half-life. One purpose for screening such combinatorial libraries is, for example, to isolate novel THAP-family or THAP domain homologs which function as either an agonist or an antagonist of the biological activities of the wild-type protein, or alternatively, possess novel activities all together. For example, mutagenesis can give rise to THAP-family homologs which have intracellular half-lives dramatically different than the corresponding wild-type protein. The altered protein can be rendered either more stable or less stable to proteolytic degradation or other cellular

process which result in destruction of, or otherwise inactivation of, a THAP-family protein. Such THAP-family homologs, and the genes which encode them, can be utilized to alter the envelope of expression for a particular recombinant THAP-family protein by modulating the half-life of the recombinant protein. For instance, a short half-life can give rise to more transient biological effects associated with a particular recombinant THAP-family protein and, when part of an inducible expression system, can allow tighter control of recombinant protein levels within a cell. As above, such proteins, and particularly their recombinant nucleic acid constructs, can be used in gene therapy protocols.

In an illustrative embodiment of this method, the amino acid sequences for a population of THAP-family homologs or other related proteins are aligned, preferably to promote the highest homology possible. Such a population of variants can include, for example, THAP-family homologs from one or more species, or THAP-family homologs from the same species but which differ due to mutation. Amino acids which appear at each position of the aligned sequences are selected to create a degenerate set of combinatorial sequences. There are many ways by which the library of potential THAP-family homologs can be generated from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be carried out in an automatic DNA synthesizer, and the synthetic genes then be ligated into an appropriate gene for expression. The purpose of a degenerate set of genes is to provide, in one mixture, all of the sequences encoding the desired set of potential THAP-family sequences. The synthesis of degenerate oligonucleotides is well known in the art (see for example. Narang, SA (1983) Tetrahedron 393; Itakura et al. (1981) Recombinant DNA, Proc 3rd Cleveland Sympos. Macromolecules, ed. AG Walton, Amsterdam: Elsevier pp. 273-289; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al. (1984) Science 198:1056; Ike et al. (1983) Nucleic Acid Res. 11:477. Such techniques have been employed in the directed evolution of other proteins (see, for example, Scott et al. (1990) Science 249:386-390; Roberts et al. (1992) PNAS 89:2429-2433; Devlin et al. (1990) Science 249: 404-406; Cwirla et al. (1990) PNAS 87: 6378-6382; as well as U.S. Patents Nos: 5,223,409, 5,198,346, and 5,096,815). The disclosures of the above references are incorporated herein by reference in their entireties.

Alternatively, other forms of mutagenesis can be utilized to generate a combinatorial library, particularly where no other naturally occurring homologs have yet been sequenced. For example, THAP-family homologs (both agonist and antagonist forms) can be generated and isolated from a library by screening using, for example, alanine scanning mutagenesis and the like (Ruf et al. (1994) Biochemistry 33:1565-1572; Wang et al. (1994) J Biol. Chem. 269:3095-3099; Balint et al. (1993) Gene 137:109-118; Grodberg et al. (1993) Eur. J Biochem. 218:597-601; Nagashima et al. (1993) J Biol. Chem. 268:2888-2892; Lowman et al. (1991) Biochemistry 30:10832-10838; and Cunningham et al. (1989) Science 244:1081-1085), by linker scanning mutagenesis (Gustin et al. (1993) Virology 193:653-660; Brown et al. (1992) Mol. Cell Biol. 12:2644 2652; McKnight et al. (1982) Science 232:316); by saturation mutagenesis (Meyers et al. (1986) Science 232:613); by PCR mutagenesis (Leung et al. (1989) Method Cell Mol Biol 1: 1-19); or by random mutagenesis (Miller et al. (1992) A Short Course in Bacterial Genetics, CSHL Press, Cold Spring Harbor, NY; and Greener et al. (1994) Strategies in Mol Biol 7:32-34, the disclosures of which are incorporated herein by reference in their entireties).

A wide range of techniques are known in the art for screening gene products of combinatorial libraries made by point mutations, as well as for screening cDNA libraries for gene products having a certain property. Such techniques will be generally adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of THAP-family proteins. The most widely used techniques for screening large gene libraries typically comprises cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates relatively easy isolation of the vector encoding the gene whose product was detected.

Each of the illustrative assays described below are amenable to high through-put analysis as necessary to screen large numbers of degenerate THAP-family or THAP domain sequences created by combinatorial mutagenesis techniques. In one screening assay, the candidate gene products are displayed on the surface of a cell or viral particle, and the ability of particular cells or viral particles to bind a THAP-family target molecule (protein or DNA) via this gene product is detected in a "panning assay". For

instance, the gene library can be cloned into the gene for a surface membrane protein of a bacterial cell, and the resulting fusion protein detected by panning (Ladner et al., WO 88/06630; Fuchs et al. (1991) BiolTechnology 9:1370-1371, and Goward et al. (1992) TIBS 18:136 140). In a similar fashion, fluorescently labeled THAP-family target can be used to score for potentially functional THAP-family homologs. Cells can be visually inspected and separated under a fluorescence microscope, or, where the morphology of the cell permits, separated by a fluorescence- activated cell sorter.

In an alternate embodiment, the gene library is expressed as a fusion protein on the surface of a viral particle. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at very high concentrations, a large number of phage can be screened at one time. Second, since each infectious phage displays the combinatorial gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical *E. coli* filamentous phages M13, fd, and fl are most often used in phage display libraries, as either of the phage gIll or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle (Ladner et al. PCT publication WO 90/02909; Garrard et al., PCT publication WO 92/09690; Marks et al. (1992) J Biol. Chem. 267:16007-16010; Griffiths et al. (1993) EMBO J 12:725-734; Clackson et al. (1991) Nature 352:624-628; and Barbas et al. (1992) PNAS 89:4457 4461, the disclosures of which are incorporated herein by reference in their entireties). In an illustrative embodiment, the recombinant phage antibody system (RPAS, Pharmacia Catalog number 27-9400-01) can be easily modified for use in expressing THAP-family combinatorial libraries, and the THAP-family phage library can be panned on immobilized THAP family target molecule (glutathione immobilized THAP-family target-GST fusion proteins or immobilized DNA). Successive rounds of phage amplification and panning can greatly enrich for THAP-family homologs which retain an ability to bind a THAP-family target and which can subsequently be screened further for biological activities in automated assays, in order to distinguish between agonists and antagonists.

The invention also provides for identification and reduction to functional minimal size of the THAP-family domains, particularly a THAP domain of the subject THAP-family to generate mimetics, e.g. peptide or non-peptide agents, which are able to disrupt binding of a polypeptide of the present invention with a THAP-family target molecule (protein or DNA). Thus, such mutagenic techniques as described above are also useful to map the determinants of THAP-family proteins which participate in protein-protein or protein-DNA interactions involved in, for example, binding to a THAP-family or THAP domain target protein or DNA. To illustrate, the critical residues of a THAP-family protein which are involved in molecular recognition of the THAP-family target can be determined and used to generate THAP-family target-13P-derived peptidomimetics that competitively inhibit binding of the THAP-family protein to the THAP-family target. By employing, for example, scanning mutagenesis to map the amino acid residues of a particular THAP-family protein involved in binding a THAP-family target, peptidomimetic compounds can be generated which mimic those residues in binding to a THAP-family target, and which, by inhibiting binding of the THAP-family protein to the THAP-family target molecule, can interfere with the function of a THAP-family protein in transcriptional regulation of one or more genes. For instance, non hydrolyzable peptide analogs of such residues can be generated using retro-inverse peptides (e.g., see U.S. Patents 5,116,947 and 5,219,089; and Pallai et al. (1983) Int J Pept Protein Res 21:84-92), benzodiazepine (e.g., see Freidinger et al. in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al. in Peptides.- Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), substituted gamma lactam rings (Garvey et al. in Peptides: Chemistry and Biology, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), keto-methylene pseudopeptides (Ewenson et al. (1986) J Med Chem 29:295; and Ewenson et al. in Peptides: Structure and Function (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985), P-turn dipeptide cores (Nagai et al. (1985) Tetrahedron Left 26:647; and Sato et al. (1986) J Chem Soc Perkin Trans 1: 123 1), and P-aminoalcohols (Gordon et al. (1985) Biochem Biophys Res Commun 126:419; and Dann et al. (1986) Biochem Biophys Res Commun 134:71, the disclosures of which are incorporated herein by reference in their entireties).

An isolated THAP-family or THAP domain protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind THAP-family or THAP domain proteins using standard techniques for polyclonal and monoclonal antibody preparation. A full-length THAP-family protein can be used or, alternatively, the invention provides antigenic peptide fragments of THAP-family or THAP domain proteins for use as immunogens. Any fragment of the THAP-family or THAP domain protein which contains at least one antigenic determinant may be used to generate antibodies. The antigenic peptide of a THAP-family or THAP domain protein comprises at least 8 amino acid residues of an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-114 and encompasses an epitope of a THAP-family or THAP domain protein such that an antibody raised against the peptide forms a specific immune complex with a THAP-family or THAP domain protein. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

Preferred epitopes encompassed by the antigenic peptide are regions of a THAP-family or THAP domain protein that are located on the surface of the protein, e.g., hydrophilic regions.

A THAP-family or THAP domain protein immunogen typically is used to prepare antibodies by immunizing a suitable subject, (e.g., rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed THAP-family or THAP domain protein or a chemically synthesized THAP-family or THAP domain polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic THAP-family or THAP domain protein preparation induces a polyclonal anti-THAP-family or THAP domain protein antibody response.

The invention concerns antibody compositions, either polyclonal or monoclonal, capable of selectively binding, or selectively bind to an epitope-containing a polypeptide comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, 100, or more than 100 amino acids of an amino acid sequence selected from the group consisting of amino

acid positions 1 to approximately 90 of SEQ ID NOs: 1-114. The invention also concerns a purified or isolated antibody capable of specifically binding to a mutated THAP-family or THAP domain protein or to a fragment or variant thereof comprising an epitope of the mutated THAP-family or THAP domain protein.

Oligomeric Forms of THAP1

Certain embodiments of the present invention encompass THAP1 polypeptides in the form of oligomers, such as dimers, trimers, or higher oligomers. Oligomers may be formed by disulfide bonds between cysteine residues on different THAP1 polypeptides, for example. In other embodiments, oligomers comprise from two to four THAP1 polypeptides joined by covalent or non-covalent interactions between peptide moieties fused to the THAP1 polypeptides. Such peptide moieties may be peptide linkers (spacers), or peptides that have the property of promoting oligomerization. Leucine zippers and certain polypeptides derived from antibodies are among the peptides that can promote oligomerization of THAP1 polypeptides attached thereto. DNA sequences encoding THAP1 oligomers, or fusion proteins that are components of such oligomers, are provided herein.

In one embodiment of the invention, oligomeric THAP1 may comprise two or more THAP1 polypeptides joined through peptide linkers. Examples include those peptide linkers described in U.S. Patent No. 5,073,627, the disclosure of which is incorporated herein by reference in its entirety. Fusion proteins comprising multiple THAP1 polypeptides separated by peptide linkers may be produced using conventional recombinant DNA technology.

Another method for preparing THAP1 oligomers involves use of a leucine zipper. Leucine zipper domains are peptides that promote oligomerization of the proteins in which they are found. Leucine zippers were originally identified in several DNA-binding proteins (Landschulz et al., Science 240:1759, 1988), and have since been found in a variety of different proteins. Among the known leucine zippers are naturally occurring peptides and derivatives thereof that dimerize or trimerize. Examples of leucine zipper domains suitable for producing THAP1 oligomers are those described International Publication WO 94/10308, the disclosure of which is incorporated herein by reference in its entirety. Recombinant fusion proteins

comprising a THAP1 polypeptide fused to a peptide that dimerizes or trimerizes in solution are expressed in suitable host cells, and the resulting soluble oligomeric THAP1 is recovered from the culture supernatant.

In some embodiments of the invention, a THAP1 dimer is created by fusing THAP1 to an Fc region polypeptide derived from an antibody, in a manner that does not substantially affect the binding of THAP1 to the chemokine SLC/CCL21. Preparation of fusion proteins comprising heterologous polypeptides fused to various portions of antibody-derived polypeptides (including Fc region) has been described, e.g., by Ashkenazi et al. (1991) *PNAS* 88:10535, Byrn et al. (1990) Nature 344:667, and Hollenbaugh and Aruffo "Construction of Immunoglobulin Fusion Proteins", in Current Protocols in Immunology, Supp. 4, pages 10.19.1 -10.19-11, 1992, the disclosures of which are incorporated herein by reference in their entireties. The THAP1/Fc fusion proteins are allowed to assemble much like antibody molecules, whereupon interchain disulfide bonds form between Fc polypeptides, yielding divalent THAP1. Similar fusion proteins of TNF receptors and Fc (see for example Moreland et al. (1997) N. Engl. J. Med. 337(3):141-147; van der Poll et al. (1997) Blood 89(10):3727-3734; and Ammann et al. (1997) J. Clin. Invest. 99(7):1699-1703) have been used successfully for treating rheumatoid arthritis. Soluble derivatives have also been made of cell surface glycoproteins in the immunoglobulin gene superfamily consisting of an extracellular domain of the cell surface glycoprotein fused to an immunoglobulin constant (Fc) region (see e.g., Capon, D. J. et al. (1989) Nature 337:525-531 and Capon U.S. Pat. Nos. 5,116,964 and 5,428,130 [CD4-IgG1 constructs]; Linsley, P. S. et al. (1991) J. Exp. Med. 173:721-730 [a CD28-IgG1 construct and a B7-1-IgG1 construct]; and Linsley, P. S. et al. (1991) J. Exp. Med. 174:561-569 and U.S. Patent No. 5,434,131 [a CTLA4-IgG1], the disclosures of which are incorporated herein by reference in their entirety). Such fusion proteins have proven useful for modulating receptor-ligand interactions.

Some embodiments relate to THAP-immunoglobulin fusion proteins and THAP SLC-binding domain fusions with immunoglobulin molecules or fragments thereof. Such fusions can be produced using standard methods, for example, by creating an expression vector encoding the SLC/CCL21 chemokine-binding protein THAP1 fused to the antibody polypeptide and inserting the vector into a suitable host cell. One

suitable Fc polypeptide is the native Fc region polypeptide derived from a human IgG1, which is described in International Publication WO 93/10151, the disclosure of which is incorporated herein by reference in its entirety. Another useful Fc polypeptide is the Fc mutein described in U.S. Patent No. 5,457,035, the disclosure of which is incorporated herein by reference in its entirety. The amino acid sequence of the mutein is identical to that of the native Fc sequence presented in International Publication WO 93/10151, the disclosure of which is incorporated herein by reference in its entirety, except that amino acid 19 has been changed from Leu to Ala, amino acid 20 has been changed from Leu to Glu, and amino acid 22 has been changed from Gly to Ala. This mutein Fc exhibits reduced affinity for immunoglobulin receptors.

SLC-binding fragments of human THAP1, rather than the full protein, can also be employed in methods of the invention. Fragments may be less immunogenic than the corresponding full-length proteins. The ability of a fragment to bind chemokine SLC can be determined using a standard assay. Fragments can be prepared by any of a number of conventional methods. For example, a desired DNA sequence can be synthesized chemically or produced by restriction endonuclease digestion of a full length cloned DNA sequence and isolated by electrophoresis on agarose gels. Linkers containing restriction endonuclease cleavage sites can be employed to insert the desired DNA fragment into an expression vector, or the fragment can be digested at naturally-present cleavage sites. The polymerase chain reaction (PCR) can also be employed to isolate a DNA sequence encoding a desired protein fragment. Oligonucleotides that define the termini of the desired fragment are used as 5' and 3' primers in the PCR procedure. Additionally, known mutagenesis techniques can be used to insert a stop codon at a desired point, e.g., immediately downstream of the codon for the last amino acid of the desired fragment.

In other embodiments, THAP1 or a biologically active fragment thereof, for example, an SLC-binding domain of THAP1 may be substituted for the variable portion of an antibody heavy or light chain. If fusion proteins are made with both heavy and light chains of an antibody, it is possible to form a THAP1 oligomer with at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, or more than nine THAP1 polypeptides.

In some embodiments of the present invention, THAP-SLC binding can be provided to decrease the biological availability of SLC or otherwise disrupt the activity of SLC. For example, THAP-family polypeptides, SLC-binding domains of THAP-family polypeptides, THAP oligomers, and SLC-binding domain-THAP1-immunoglobulin fusion proteins of the invention can be used to interact with SLC thereby preventing it from performing its normal biological role. In some embodiments, the entire THAP1 polypeptide (SEQ ID NO: 3) can be used to bind to SLC. In other embodiments, fragments of THAP1, such as the SLC-binding domain of the THAP1 (amino acids 143-213 of SEQ ID NO: 3) can used to bind to SLC. Such fragments can be from at least 8, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 160, at least 170, at least 180, at least 190, at least 200, at least 210 or at least 213 consecutive amino acids of SEQ ID NO: 3. In some embodiments, fragments can be from at least 8, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65 or at least 70 consecutive amino acids of (amino acids 143-213 of SEQ ID NO: 3). THAP-family polypeptides that may be capable of binding SLC, for example THAP2-11 and THAP0 or biologically active fragments thereof can also be used to bind to SLC so as to decrease its biological availability or otherwise disrupt the activity of this chemokine.

In some embodiments, a plurality of THAP-family proteins, such as a fusion of two or more THAP1 proteins or fragments thereof which comprise an SLC-binding domain (amino acids 143-213 of SEQ ID NO: 3) can be used to bind SLC. For example, oligomers comprising THAP1 fragments of a size of at least 8, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65 or at least 70 consecutive amino acids of SEQ ID NO: 3 (amino acids 143-213) can be generated. Amino acid fragments which make up the THAP oligomer may be of the same or different lengths. In some embodiments, the entire THAP1 protein or biologically active portions thereof may be fused together to form an oligomer capable of binding to SLC. THAP-family polypeptides that may be capable of binding SLC, for example THAP2-11 and THAP0,

the THAP-family polypeptides of SEQ ID NOs: 16-114 or biologically active fragments thereof can also be used to create oligomers which bind to SLC so as to decrease its biological availability or otherwise disrupt the activity of this chemokine.

According to another embodiment of the present invention, THAP-family proteins, such as THAP1 or portion of THAP1 which comprise an SLC binding domain (amino acids 143-213 of SEQ ID NO: 3), may be fused to an immunoglubulin or portion thereof. The portion may be an entire immunoglobulin, such as IgG, IgM, IgA or IgE. Additionally, portions of immunoglobulins, such as an Fc domain of the immunoglobulin, can be fused to a THAP-family polypeptide, such as THAP1, fragments thereof or oligomers thereof . Fragments of THAP1 can be, for example, at least 8, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65 or at least 70 consecutive amino acids of SEQ ID NO: 3 (amino acids 143-213). In some embodiments, THAP-family polypeptides that may be capable of binding SLC, for example THAP2-11 and THAP0, the THAP-family polypeptides of SEQ ID NOs: 16-114 or biologically active fragments thereof can also be used to form immunglobulin fusion that bind to SLC so as to decrease its biological availability or otherwise disrupt the activity of this chemokine.

In accordance with another aspect of the invention, THAP-family polypeptides, SLC-binding domains of THAP-family polypeptides, THAP oligomers, and SLC-binding domain-THAP1-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions. Such pharmaceutical compositions can be used to decrease the bioavailability and functionality of SLC. For example, THAP-family polypeptides, SLC-binding domains of THAP-family polypeptides, THAP oligomers, and SLC- binding domain-THAP1-immunoglobulin fusion proteins of the present invention can be administered to a subject to inhibit an interaction between SLC and its receptor, such as CCR7, on the surface of cells, to thereby suppress SLC-mediated responses. The inhibition of chemokine SLC may be useful therapeutically for both the treatment of inflammatory or proliferative disorders, as well as modulating (e.g., promoting or inhibiting) cell differentiation, cell proliferation, and/or cell death.

In an additional embodiment of the present invention, the THAP-family polypeptides, SLC-binding domains of THAP-family polypeptides, THAP oligomers,

and SLC- binding domain-THAP1-immunoglobulin fusion proteins of the present invention can be used to detect the presence of SLC in a biological sample and in screening assays to identify molecules which inhibit the interaction of THAP1 with SLC. Such screening assays are similar to those described below for PAR4-THAP interactions.

Certain aspects of the present invention related to a method of identifying a test compound that modulates THAP-mediated activites. In some cases the THAP-mediated acitivity is SLC-binding. Test compounds which affect THAP-SLC binding can be identified using a screening method wherein a THAP-family polypeptide or a biologically active fragment thereof is contacted with a test compound. In some embodiments, the THAP-family polypeptide comprises an amino acid sequence having at least 30% amino acid identity to an amino acid sequence of SEQ ID NO: 1 or SEQ ID NO: 2. Whether the test compound modulates the binding of SLC with a THAP-family polypeptide, such as THAP1 (SEQ ID NO: 3), is determined by determining whether the test compound modulates the activity of the THAP-family polypeptide or biologically active fragment thereof. Biologically active framents of a THAP-family polypeptide may be at least 5, at least 8, at least 10, at least 12, at least 15, at least 18, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 160, at least 170, at least 180, at least 190, at least 200, at least 210, at least 220 or at least more than 220 amino acids in length. A determination that the test compound modulates the activity of said polypeptide indicates that the test compound is a candidate modulator of THAP-mediated activities.

Although THAP-family polypeptides, SLC-binding domains of THAP-family polypeptides, THAP oligomers, and SLC-binding domain-THAP1-immunoglobulin fusion proteins can be used for the above-mentioned SLC interactions, it will be appreciated that homologs of THAP-family polypeptides, SLC-binding domains of THAP-family polypeptides, THAP oligomers, and SLC-binding domain-THAP1-immunoglobulin fusion proteins can be used in place of THAP-family polypeptides, SLC-binding domains of THAP-family polypeptides, THAP oligomers, and SLC-binding domain-THAP1-immunoglobulin fusion proteins. For example, homologs having at least about 30-40% identity, preferably at least about 40-50% identity, more

preferably at least about 50-60%, and even more preferably at least about 60-70%, 70-80%, 80%, 90%, 95%, 97%, 98%, 99% or 99.8% identity across the amino acid sequences of SEQ ID NOs: 1-114 or portions thereof can be used.

Although this section, entitled "Oligomeric Forms of THAP-1," describes THAP-family polypeptides, SLC-binding domains of THAP-family polypeptides, THAP oligomers, SLC-binding domain-THAP1-immunoglobulin fusion proteins and homologs of these polypeptides as well as methods of using such polypeptides, it will be appreciated that such polypeptides are included in the class of THAP-type chemokine-binding agents. Accordingly, the above description also applies to THAP-type chemokine-binding agents. It will be appreciated that THAP-type chemokine-binding agents will be used for applications which include, but are not limited to, chemokine binding, inhibiting or enhancing chemokine activity, chemokine detection, reducing the symptoms associated with a chemokine influenced or mediated condition, and reducing or preventing inflammation or other chemokine mediated conditions. THAP-type chemokine-binding agents can also be used in the kits, devices, compositions, and procedures described elsewhere herein.

In some embodiments of the present invention, THAP-type chemokine-binding agents bind to or otherwise modulate the activity of one or more chemokines selected from the group consisting of XCL1, XCL2, CCL1, CCL2, CCL3, CCL3L1, SCYA3L2, CCL4, CCL4L, CCL5, CCL6, CCL7, CCL8, SCYA9, SCYA10, CCL11, SCYA12, CCL13, CCL14, CCL15, CCL16, CCL17, CCL18, CCL19, CCL20, CCL21, CCL22, CCL23, CCL24, CCL25, CCL26, CCL27, CCL28, clone 391, CARP CC-1, CCL1, CK-1, regakine-1, K203, CXCL1, CXCL1P, CXCL2, CXCL3, PF4, PF4V1, CXCL5, CXCL6, PPBP, SPBPBP, IL8, CXCL9, CXCL10, CXCL11, CXCL12, CXCL14, CXCL15, CXCL16, NAP-4, LFCA-1, Scyba, JSC, VHSV-induced protein, CX3CL1, and fCL1.

Primers and probes

Primers and probes of the invention can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphodiester method of Narang SA et al (Methods Enzymol 1979;68:90-98), the phosphodiester method of Brown EL et al

(Methods Enzymol 1979;68:109-151), the diethylphosphoramidite method of Beaucage et al (Tetrahedron Lett 1981, 22: 1859-1862) and the solid support method described in EP 0 707 592, the disclosures of which are incorporated herein by reference in their entireties.

Detection probes are generally nucleic acid sequences or uncharged nucleic acid analogs such as, for example peptide nucleic acids which are disclosed in International Patent Application WO 92/20702, morpholino analogs which are described in U.S. Patents Numbered 5,185,444; 5,034,506 and 5,142,047. If desired, the probe may be rendered "non-extendable" in that additional dNTPs cannot be added to the probe. In and of themselves analogs usually are non-extendable and nucleic acid probes can be rendered non-extendable by modifying the 3' end of the probe such that the hydroxyl group is no longer capable of participating in elongation. For example, the 3' end of the probe can be functionalized with the capture or detection label to thereby consume or otherwise block the hydroxyl group.

Any of the polynucleotides of the present invention can be labeled, if desired, by incorporating any label known in the art to be detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive substances (including, ^{32}P, ^{35}S, ^{3}H, ^{125}I), fluorescent dyes (including, 5-bromodesoxyuridin, fluorescein, acetylaminofluorene, digoxigenin) or biotin. Preferably, polynucleotides are labeled at their 3' and 5' ends. Examples of non-radioactive labeling of nucleic acid fragments are described in (Urdea et al. (Nucleic Acids Research. 11:4937-4957, 1988) or Sanchez-Pescador et al. (J. Clin. Microbiol. 26(10):1934-1938, 1988). In addition, the probes according to the present invention may have structural characteristics such that they allow the signal amplification, such structural characteristics being, for example, branched DNA probes as those described by Urdea et al (Nucleic Acids Symp. Ser. 24:197-200, 1991) or in the European patent No. EP 0 225 807 (Chiron).

A label can also be used to capture the primer, so as to facilitate the immobilization of either the primer or a primer extension product, such as amplified DNA, on a solid support. A capture label is attached to the primers or probes and can be a specific binding member which forms a binding pair with the solid's phase reagent's specific binding member (e.g. biotin and streptavidin). Therefore depending

upon the type of label carried by a polynucleotide or a probe, it may be employed to capture or to detect the target DNA. Further, it will be understood that the polynucleotides, primers or probes provided herein, may, themselves, serve as the capture label. For example, in the case where a solid phase reagent's binding member is a nucleic acid sequence, it may be selected such that it binds a complementary portion of a primer or probe to thereby immobilize the primer or probe to the solid phase. In cases where a polynucleotide probe itself serves as the binding member, those skilled in the art will recognize that the probe will contain a sequence or "tail" that is not complementary to the target. In the case where a polynucleotide primer itself serves as the capture label, at least a portion of the primer will be free to hybridize with a nucleic acid on a solid phase. DNA labeling techniques are well known to the skilled technician.

The probes of the present invention are useful for a number of purposes. They can be notably used in Southern hybridization to genomic DNA. The probes can also be used to detect PCR amplification products. They may also be used to detect mismatches in a THAP-family gene or mRNA using other techniques.

Any of the nucleic acids, polynucleotides, primers and probes of the present invention can be conveniently immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, nitrocellulose strips, membranes, microparticles such as latex particles, sheep (or other animal) red blood cells, duracytes and others. The solid support is not critical and can be selected by one skilled in the art. Thus, latex particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and duracytes are all suitable examples. Suitable methods for immobilizing nucleic acids on solid phases include ionic, hydrophobic, covalent interactions and the like. A solid support, as used herein, refers to any material which is insoluble, or can be made insoluble by a subsequent reaction. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent. Alternatively, the solid phase can retain an additional receptor which has the ability to attract and immobilize the capture reagent. The additional receptor can include a charged substance that is oppositely charged with respect to the capture reagent itself or

to a charged substance conjugated to the capture reagent. As yet another alternative, the receptor molecule can be any specific binding member which is immobilized upon (attached to) the solid support and which has the ability to immobilize the capture reagent through a specific binding reaction. The receptor molecule enables the indirect binding of the capture reagent to a solid support material before the performance of the assay or during the performance of the assay. The solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cells, duracytes and other configurations known to those of ordinary skill in the art. The nucleic acids, polynucleotides, primers and probes of the invention can be attached to or immobilized on a solid support individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the invention to a single solid support. In addition, polynucleotides other than those of the invention may be attached to the same solid support as one or more polynucleotides of the invention.

Any polynucleotide provided herein may be attached in overlapping areas or at random locations on a solid support. Alternatively the polynucleotides of the invention may be attached in an ordered array wherein each polynucleotide is attached to a distinct region of the solid support which does not overlap with the attachment site of any other polynucleotide. Preferably, such an ordered array of polynucleotides is designed to be "addressable" where the distinct locations are recorded and can be accessed as part of an assay procedure. Addressable polynucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. The knowledge of the precise location of each polynucleotides location makes these "addressable" arrays particularly useful in hybridization assays. Any addressable array technology known in the art can be employed with the polynucleotides of the invention. One particular embodiment of these polynucleotide arrays is known as the Genechips, and has been generally described in US Patent 5,143,854; PCT publications WO 90/15070 and 92/10092, the disclosures of which are incorporated herein by reference in their entireties.

Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof.

Vectors may have particular use in the preparation of a recombinant protein of the invention, or for use in gene therapy. Gene therapy presents a means to deliver a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof to a subject in order to regulate apoptosis for treatment of a disorder.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a THAP-family or THAP domain nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the

nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990), the disclosure of which is incorporated herein by reference in its entirety. Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., THAP-family proteins, mutant forms of THAP-family proteins, fusion proteins, or fragments of any of the preceding proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof in prokaryotic or eukaryotic cells. For example, THAP-family or THAP domain proteins can be expressed in bacterial cells such as E. coli, insect cells (using baculovirus expression vectors) yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990), the disclosure of which is incorporated herein by reference in its entirety. Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in E. coli with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein;

2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D. B. and Johnson, K. S. (1988) Gene 67:31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.), the disclosures of which are incorporated herein by reference in their entireties, which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Purified fusion proteins can be utilized in THAP-family activity assays, (e.g., direct assays or competitive assays described in detail below), or to generate antibodies specific for THAP-family or THAP domain proteins, for example. In a preferred embodiment, a THAP-family or THAP domain fusion protein expressed in a retroviral expression vector of the present invention can be utilized to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (for example, six (6) weeks).

Examples of suitable inducible non-fusion E. coli expression vectors include pTrc (Amann et al., (1988) Gene 69:301-315) and pET 11d (Studier et al., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 60-89), the disclosures of which are incorporated herein by reference in their entireties. Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn 1). This viral polymerase is supplied by host strains BL21 (DE3) or HMS174(DE3) from a resident prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in E. coli is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the

recombinant protein (Gottesman, S., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 119-128, the disclosure of which is incorporated herein by reference in its entirety). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in E. coli (Wada et al., (1992) Nucleic Acids Res. 20:2111-2118, the disclosure of which is incorporated herein by reference in its entirety). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the THAP-family expression vector is a yeast expression vector. Examples of vectors for expression in yeast S. cerivisae include pYepSec 1 (Baldari, et al., (1987) Embo J. 6:229-234), pMFa (Kurjan and Herskowitz, (1982) Cell 30:933-943), pJRY88 (Schultz et al., (1987) Gene 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.), the disclosures of which are incorporated herein by reference in their entireties.

Alternatively, THAP-family or THAP domain proteins can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf9 cells) include the pAc series (Smith et al. (1983) Mol. Cell Biol. 3:2156-2165) and the pVL series (Lucklow and Summers (1989) Virology 170:31-39), the disclosures of which are incorporated herein by reference in their entireties. In particularly preferred embodiments, THAP-family proteins are expressed according to Karniski et al, Am. J. Physiol. (1998) 275: F79-87, the disclosure of which is incorporated herein by reference in its entirety.

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) Nature 329:840) and pMT2PC (Kaufman et al. (1987) EMBO J. 6:187-195), the disclosures of which are incorporated herein by reference in their entireties. When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,

Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989, the disclosure of which is incorporated herein by reference in its entirety. In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art, and are further described below.

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to THAP-family mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. et al., Antisense RNA as a molecular tool for genetic analysis, Reviews--Trends in Genetics, Vol. 1(1) 1986, the disclosure of which is incorporated herein by reference in its entirety.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such term refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a THAP-family protein can be expressed in bacterial cells such as E. coli, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells or human

cells). Other suitable host cells are known to those skilled in the art, including mouse 3T3 cells as further described in the Examples.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989, the disclosure of which is incorporated herein by reference in its entirety), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a THAP-family protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) a THAP-family protein. Accordingly, the invention further provides methods for producing a THAP-family protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a THAP-family protein has been introduced) in a suitable medium such that a THAP-family protein is produced. In another embodiment, the method further comprises isolating a THAP-family protein from the medium or the host cell.

In another embodiment, the invention encompassesa method comprising: providing a cell capable of expressing a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof, culturing said cell in a suitable medium such that a THAP-family or THAP domain protein is produced, and isolating or purifying the THAP-family or THAP domain protein from the medium or cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals, such as for the study of disorders in which THAP family proteins are implicated. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which THAP-family- or THAP domain- coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous THAP-family or THAP domain sequences have been introduced into their genome or homologous recombinant animals in which endogenous THAP-family or THAP domain sequences have been altered. Such animals are useful for studying the function and/or activity of a THAP-family or THAP domain polypeptide or fragment thereof and for identifying and/or evaluating modulators of a THAP-family or THAP domain activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous THAP-family or THAP domain gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Pat. Nos. 4,736,866 and 4,870,009, both by Leder et al., U.S. Pat. No. 4,873,191 by Wagner et al. and in Hogan, B., Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory

Press, Cold Spring Harbor, N.Y., 1986, the disclosures of which are incorporated herein by reference in their entireties).

Gene Therapy Vectors

Prefered vectors for administration to a subject can be constructed according to well known methods. Vectors will comprise regulatory elements (e.g. promotor, enhancer, etc) capable of directing the expression of the nucleic acid in the targeted cell. Thus, where a human cell is targeted, it is preferable to position the nucleic acid coding region adjacent to and under the control of a promoter that is capable of being expressed in a human cell.

In various embodiments, the human cytornegalovirus (CMV) immediate early gene promoter, the SV40 early promoter, the Rous sarcoma virus long terminal repeat, P actin, rat insulin promoter and glyceraldehyde-3 -phosphate dehydrogenase can be used to obtain high-level expression of the coding sequence of interest. The use of other viral or mammalian cellular or bacterial phage promoters which are well-known in the art to achieve expression of a coding sequence of interest is contemplated as well, provided that the levels of expression are sufficient for a given purpose. By employing a promoter with well-known properties, the level and pattern of expression of the protein of interest following transfection or transformation can be optimized.

Selection of a promoter that is regulated in response to specific physiologic or synthetic signals can permit inducible expression of the gene product. For example in the case where expression of a transgene, or transgenes when a multicistronic vector is utilized, is toxic to the cells in which the vector is produced in, it may be desirable to prohibit or reduce expression of one or more of the transgenes. Several inducible promoter systems are available for production of viral vectors where the transgene product may be toxic.

The ecdysone system (Invitrogen, Carlsbad, CA) is one such system. This system is designed to allow regulated expression of a gene of interest in mammalian cells. It consists of a tightly regulated expression mechanism that allows virtually no basal level expression of the transgene, but over 200-fold inducibility. The system is based on the heterodimeric ecdysone receptor of Drosophila, and when ecdysone or an analog such as muristerone A binds to the receptor, the receptor activates a promoter to

turn on expression of the downstream transgene high levels of mRNA transcripts are attained. In this system, both monomers of the heterodimeric receptor are constituitively expressed from one vector, whereas the ecdysone-responsive promoter which drives expression of the gene of interest is on another plasmid. Engineering of this type of system into the gene transfer vector of interest would therefore be useful. Cotransfection of plasmids containing the gene of interest and the receptor monomers in the producer cell line would then allow for the production of the gene transfer vector without expression of a potentially toxic transgene. At the appropriate time, expression of the transgene could be activated with ecdysone or muristeron A. Another inducible system that would be useful is the Tet-Off or Tet On system (Clontech, Palo Alto, CA) originally developed by Gossen and Bujard (Gossen and Bujard, 1992; Gossen et al, 1995). This system also allows high levels of gene expression to be regulated in response to tetracycline or tetracycline derivatives such as doxycycline. In the Tet-On system, gene expression is turned on in the presence of doxycycline, whereas in the Tet-Off system, gene expression is turned on in the absence of doxycycline. These systems are based on two regulatory elements derived from the tetracycline resistance operon of E. coli. The tetracycline operator sequence to which the tetracycline repressor binds, and the tetracycline repressor protein. The gene of interest is cloned into a plasmid behind a promoter that has tetracycline-responsive elements present in it. A second plasmid contains a regulatory element called the tetracycline-controlled transactivator, which is composed, in the Tet Off system, of the VP16 domain from the herpes simplex virus and the wild-type tertracycline repressor.

Thus in the absence of doxycycline, transcription is constituitively on. In the Tet-OnTm system, the tetracycline repressor is not wild-type and in the presence of doxycycline activates transcription. For gene therapy vector production, the Tet Off system would be preferable so that the producer cells could be grown in the presence of tetracycline or doxycycline and prevent expression of a potentially toxic transgene, but when the vector is introduced to the patient, the gene expression would be constituitively on.

In some circumstances, it may be desirable to regulate expression of a transgene in a gene therapy vector. For example, different viral promoters with varying strengths of activity may be utilized depending on the level of expression desired. In mammalian

cells, the CMV immediate early promoter if often used to provide strong transcriptional activation. Modified versions of the CMV promoter that are less potent have also been used when reduced levels of expression of the transgene are desired. When expression of a transgene in hematopoetic_cells is desired, retroviral promoters such as the LTRs from MLV or MMTV are often used. Other viral promoters that may be used depending on the desired effect include SV40, RSV LTR, HIV-1 and HfV-2 LTR, adenovirus promoters such as from the EIA, E2A, or MLP region, AAV LTR, cauliflower mosaic virus, HSV-TK, and avian sarcoma virus.

Similarly tissue specific promoters may be used to effect transcription in specific tissues or cells so as to reduce potential toxicity or undesirable effects to non-targeted tissues. For example, promoters such as the PSA, probasin, prostatic acid phosphatase or prostate-specific glandular kallikrein (hK2) may be used to target gene expression in the prostate. Similarly, promoters as follows may be used to target gene expression in other tissues.

Tissue specific promoters include in (a) *pancreas:* insulin, elastin, amylase, pdr-I, pdx-I, glucokinase; (b) *liver:* albumin PEPCK, HBV enhancer, alpha fetoprotein, apolipoprotein C, alpha-I antitrypsin, vitellogenin, NF-AB, Transthyretin; (c) *skeletal muscle:* myosin H chain, muscle creatine kinase, dystrophin, calpain p94, skeletal alpha-actin, fast troponin 1; (d) *skin:* keratin K6, keratin KI; (e) *lung*: CFTR, human cytokeratin IS (K 18), pulmonary surfactant proteins A, B and C, CC-10, Pi; (f) *smooth muscle:* sm22 alpha, SM-alpha-actin; (g) *endothelium*: endothelin- I, E-selectin, von Willebrand factor, TIE (Korhonen et al., 1995), KDR/flk-I; (h) *melanocytes*: tyrosinase; (i) *adipose tissue:* lipoprotein lipase (Zechner et al., 1988), adipsin (Spiegelman et al., 1989), acetyl-CoA carboxylase (Pape and Kim, 1989), glycerophosphate dehydrogenase (Dani et al., 1989), adipocyte P2 (Hunt et al., 1986); and (j) *blood:* P-globin.

In certain indications, it may be desirable to activate transcription at specific times after administration of the gene therapy vector. This may be done with such promoters as those that are hormone or cytokine regulatable. For example in gene therapy applications where the indication is in a gonadal tissue where specific steroids are produced or routed to, use of androgen or estrogen regulated promoters may be advantageous. Such promoters that are hormone regulatable include MMTV, MT-1, ecdysone and RuBisco. Other hormone regulated promoters such as those responsive to

thyroid, pituitary and adrenal hormones are expected to be useful in the present invention. Cytokine and inflammatory protein responsive promoters that could be used include K and T Kininogen (Kageyama et al., 1987), c-fos, TNF-alpha, C-reactive protein (Arcone et al., 1988), haptoglobin (Oliviero et al., 1987), serum amyloid A2, C/EBP alpha, IL-1, IL-6 (Poli and Cortese, 1989), Complement C3 (Wilson et al., 1990), IL-8, alpha-1 acid glycoprotein (Prowse and Baumann, 1988), alpha-1 antitypsin, lipoprotein lipase (Zechner et al., 1988), angiotensinogen (Ron et al., 1991), fibrinogen, c-jun (inducible by phorbol esters, TNF alpha, UV radiation, retinoic acid, and hydrogen peroxide), collagenase (induced by phorbol esters and retinoic acid), metallothionein (heavy metal and glucocorticoid inducible), Stromelysin (inducible by phorbol ester, interleukin-1 and EGF), alpha-2 macroglobulin and alpha- I antichymotrypsin.

It is envisioned that cell cycle regulatable promoters may be useful in the present invention. For example, in a bi-cistronic gene therapy vector, use of a strong CMV promoter to drive expression of a first gene such as p16 that arrests cells in the G1 phase could be followed by expression of a second gene such as p53 under the control of a promoter that is active in the G1 phase of the cell cycle, thus providing a "second hit" that would push the cell into apoptosis. Other promoters such as those of various cyclins, PCNA, galectin-3, E2FI, p53 and BRCAI could be used.

Tumor specific promoters such as osteocalcin, hypoxia-responsive element (HRE), NIAGE-4, CEA, alpha-fetoprotein, GRP78/BiP and tyrosinase also may be used to regulate gene expression in tumor cells. Other promoters that could be used according to the present invention include Lac-regulatable, chemotherapy inducible (e.g. MDR), and heat (hyperthermia) inducible promoters, Radiation-inducible (e.g., EGR (Joki et al., 1995)), Alpha-inhibin, RNA pol III tRNA met and other amino acid promoters, U1 snRNA (Bartlett et al., 1996), MC-1, PGK, -actin and alpha-globin. Many other promoters that may be useful are listed in Walther and Stein (1996), the disclosure of which is incorporated herein by reference.

It is envisioned that any of the above promoters alone or in combination with another may be useful according to the present invention depending on the action desired.

In addition, this list of promoters should not be considered to be exhaustive or limiting, those of skill in the art will know of other promoters that may be used in conjunction with the THAP-family and THAP domain nucleic acids and methods disclosed herein.

Enhancers

Enhancers are genetic elements that increase transcription from a promoter located at a distant position on the same molecule of DNA. Enhancers are organized much like promoters. That is, they are composed of many individual elements, each of which binds to one or more transcriptional proteins. The basic distinction between enhancers and promoters is operational. An enhancer region as a whole must be able to stimulate transcription at a distance; this need not be true of a promoter region or its component elements. On the other hand, a promoter must have one or more elements that direct initiation of RNA synthesis at a particular site and in a particular orientation, whereas enhancers lack these specificities. Promoters and enhancers are often overlapping and contiguous, often seeming to have a very similar modular organization.

Below is a list of promoters additional to the tissue specific promoters listed above, cellular promoters/enhancers and inducible promoters/enhancers that could be used in combination with the nucleic acid encoding a gene of interest in an expression construct (list of enhancers, and Table 1). Additionally, any promoter/enhancer combination (as per the Eukaryotic Promoter Data Base EPDB) could also be used to drive expression of the gene. Eukaryotic cells can support cytoplasmic transcription from certain bacterial promoters if the appropriate bacterial polymerase is provided, either as part of the delivery complex or as an additional genetic expression construct.

Suitable enhancers include: Immunoglobulin Heavy Chain; Immunoglobulin Light Chain; T-Cell Receptor; HLA DQ (x and DQ beta; beta-Interferon; Interleukin-2; Interleukin-2 Receptor; MHC Class II 5; MHC Class II HLA-DRalpha; beta-Actin; Muscle Creatine Kinase; Prealburnin (Transthyretin); Elastase I; Metallothionein; Collagenase; Albumin Gene; alpha-Fetoprotein; -Globin; beta-Globin; e-fos; c-HA-ras; Insulin; Neural Cell Adhesion Molecule (NCAM); alpha a1-Antitrypsin; H2B (TH2B) Histone; Mouse or Type I Collagen; Glucose-Regulated Proteins (GRP94 and GRP78); Rat Growth Hormone; Human Serum Amyloid A (SAA); Troponin I (TN 1); Platelet-

Derived Growth Factor; Duchenne Muscular Dystrophy; SV40; Polyoma; Retroviruses; THAPilloma Virus; Hepatitis B Virus; Human Immunodeficiency Virus; Cytomegalovirus; and Gibbon Ape Leukemia Virus.

TABLE 1

Element	Inducer
MT 11	Phorbol Ester (TPA)
Heavy metals MMTV (mouse mammary tumor Glucocorticoids virus)	
B-Interferon	poly(rI)X; poly(rc)
Adenovirus 5 E2	Ela
c-jun	Phorbol Ester (TPA), H2O2
H202 Collagenase	Phorbol Ester (TPA)
Stromelysin	Phorbol Ester (TPA), IL-1
SV40	Phorbol Ester (TPA)
Murine MX Gene	Interferon, Newcastle Disease Virus
GRP78 Gene	A23187
oc-2-Macroglobulin	IL-6
Vimentin Serum NMC Class I Gene H-2kB	Interferon
HSP70	Ela, SV40 Large T Antigen
Insulin E Box	Glucose
Proliferin	Phorbol Ester-TPA
Tumor Necrosis Factor	FMA
Thyroid Stimulating Hormone alpha Gene	Thyroid Hormone

In preferred embodiments of the invention, the expression construct comprises a virus or engineered construct derived from a viral genome. The ability of certain viruses to enter cells via receptor-mediated endocytosis and to integrate into host cell genome and express viral genes stably and efficiently have made them attractive candidates for the transfer of foreign genes into mammalian cells (Ridgeway, 1988; Nicolas and Rubenstein, 1988; Baichwal and Sugden, 1986; Temin, 1986, the disclosures of which are incorporated herein by reference). The first viruses used as gene vectors were DNA viruses including the papovaviruses (simian virus 40, bovine papilloma virus, and polyoma) (Ridgeway, 1988; Baichwal and Sugden, 1986) and adenoviruses (Ridgeway, 1988; Baichwal and Sugden, 1986). These have a relatively low capacity for foreign DNA sequences and have a restricted host spectrum.

Furthermore, their oncogenic potential and cytopathic effects in permissive cells raise safety concerns. They can accommodate only up to 8 kB of foreign genetic material but can be readily introduced in a variety of cell lines and laboratory animals (Nicolas and Rubenstein, 1988; Temin, 1986).

(iii) Polyadenylation Signals

Where a cDNA insert is employed, one will typically desire to include a polyadenylation signal to effect proper polyadenylation of the gene transcript. The nature of the polyadenylation signal is not believed to be crucial to the successful practice of the invention, and any such sequence may be employed such as human or bovine growth hormone and SV40 polyadenylation signals. Also contemplated as an element of the expression cassette is a terminator. These elements can serve to enhance message levels and to minimize read through from the cassette into other sequences.

Antisense Constructs

The term "antisense nucleic acid" is intended to refer to the oligonucleotides complementary to the base sequences of DNA and RNA. Antisense oligonucleotides, when introduced into a target cell, specifically bind to their target nucleic acid and interfere with transcription, RNA processing, transport and/or translation. Targeting double-stranded (ds) DNA with oligonucleotide leads to triple-helix formation; targeting RNA will lead to double-helix formation.

Antisense constructs may be designed to bind to the promoter and other control regions, exons, introns or even exon-intron boundaries of a gene. Antisense RNA constructs, or DNA encoding such antisense RNAs, may be employed to inhibit gene transcription or translation or both within a host cell, either in vitro or in vivo, such as within a host animal, including a human subject. Nucleic acid sequences comprising complementary nucleotides" are those which are capable of base-pairing according to the standard Watson-Crick complementary rules. That is, that the larger purines will base pair with the smaller pyrimidines to form only combinations of guanine paired with cytosine (G:C) and adenine paired with either thymine (A:T), in the case of DNA, or adenine paired with uracil (A:U) in the case of RNA.

As used herein, the terms "complementary" or "antisense sequences" mean nucleic acid sequences that are substantially complementary over their entire length and have very few base mismatches. For example, micleic acid sequences of fifteen bases in length may be termed complementary when they have a complementary nucleotide at thirteen or fourteen positions with only single or double mismatches. Naturally, nucleic

acid sequences which are "completely complementary" will be nuleic acid sequences which are entirely complementary throughout their entire length and have no base mismatches.

While all or part of the gene sequence may be employed in the context of antisense construction, statistically, any sequence 17 bases long should occur only once in the human genome and, therefore, suffice to specify a unique target sequence.

Although shorter oligomers are easier to make and increase in vivo accessibility, numerous other factors are involved in determining the specificity of hybridization. Both binding affinity and sequence specificity of an oligonucleotide to its complementary target increases with increasing length. It is contemplated that oligonucleotides of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more base pairs will be used. One can readily determine whether a given antisense nucleic acid is effective at targeting of the corresponding host cell gene simply by testing the constructs in vitro to determine whether the endogenous gene's function is affected or whether the expression of related genes having complementary sequences is affected.

In certain embodiments, one may wish to employ antisense constructs which include other elements, for example, those which include C-5 propyne pyrimidines.

Oligonucleotides which contain C-5 propyne analogues of uridine and cytidine have been shown to bind RNA with high affinity and to be potent antisense inhibitors of gene expression (Wagner et al, 1993).

Ribozyme Constructs

As an alternative to targeted antisense delivery, targeted ribozymes may be used. The term "ribozyme" refers to an RNA-based enzyme capable of targeting and cleaving particular base sequences in oncogene DNA and RNA. Ribozymes either can be targeted directly to cells, in the form of RNA oligo-nucleotides incorporating ribozyme sequences, or introduced into the cell as an expression construct encoding the desired ribozymal RNA. Ribozymes may be used and applied in much the same way as described for antisense nucleic acids.

Methods of Gene Transfer

In order to mediate the effect of transgene expression in a cell, it will be necessary to transfer the therapeutic expression constructs of the present invention into a cell. This section provides a discussion of methods and compositions of viral production and viral gene transfer, as well as non-viral gene transfer methods.

(i) Viral Vector-Mediated Transfer

The THAP-family gene is incorporated into a viral infectious particle to mediate gene transfer to a cell. Additional expression constructs encoding other therapeutic agents as described herein may also be transferred via viral transduction using infectious viral particles, for example, by transformation with an adenovirus vector of the present invention as described herein below. Alternatively, retroviral or bovine papilloma virus may be employed, both of which permit permanent transformation of a host cell with a gene(s) of interest. Thus, in one example, viral infection of cells is used in order to deliver therapeutically significant genes to a cell. Typically, the virus simply will be exposed to the appropriate host cell under physiologic conditions, permitting uptake of the virus. Though adenovirus is exemplified, the present methods may be advantageously employed with other viral or non-viral vectors, as discussed below.

Adenovirus

Adenovirus is particularly suitable for use as a gene transfer vector because of its mid-sized DNA genome, ease of manipulation, high titer, wide target-cell range, and high infectivity. The roughly 36 kB viral genome is bounded by 100-200 base pair (bp) inverted terminal repeats (ITR), in which are contained cis acting elements necessary for viral DNA replication and packaging. The early (E) and late (L) regions of the genome that contain different transcription units are divided by the onset of viral DNA replication.

The El region (EIA and EIB) encodes proteins responsible for the regulation of transcription of the viral genome and a few cellular genes. The expression of the E2 region (E2A and E2B) results in the synthesis of the proteins for viral DNA replication.

These proteins are involved in DNA replication, late gene expression, and host cell shut off (Renan, 1990). The products of the late genes (L I, L2, U, L4 and L5),

including the majority of the viral capsid proteins, are expressed only after significant processing of a single primary transcript issued by the major late promoter (MLP). The MLP (located at 16.8 map units) is particularly efficient during the late phase of infection, and all the mRNAs issued from this promoter possess a 5' tripartite leader (TL) sequence which makes them preferred mRNAs for translation.

In order for adenovirus to be optimized for gene therapy, it is necessary to maximize the carrying capacity so that large segments of DNA can be included. It also is very desirable to reduce the toxicity and immunologic reaction associated with certain adenoviral products. The two goals are, to an extent, coterminous in that elimination of adenoviral genes serves both ends. By practice of the present invention, it is possible achieve both these goals while retaining the ability to manipulate the therapeutic constructs with relative case.

The large displacement of DNA is possible because the cis elements required for viral DNA replication all are localized in the inverted terminal repeats (ITR) (100-200 bp) at either end of the linear viral genome. Plasmids containing ITR's can replicate in the presence of a non-defective adenovirus (Hay et al., 1984). Therefore, inclusion of these elements in an adenoviral vector should permit replication.

In addition, the packaging signal for viral encapsidation is localized between 194 385 bp (0.5-1.1 map units) at the left end of the viral genome (Hearing et al., 1987). This signal mimics the protein recognition site in bacteriophage k DNA where a specific sequence close to the left end, but outside the cohesive end sequence, mediates the binding to proteins that are required for insertion of the DNA into the head structure. El substitution vectors of Ad have demonstrated that a 450 bp (0-1.25 map units) fragment at the left end of the viral genome could direct packaging in 293 cells (Levrero et al., 1991).

Previously, it has been shown that certain regions of the adenoviral genome can be incorporated into the genome of mammalian cells and the genes encoded thereby expressed. These cell lines are capable of supporting the replication of an adenoviral vector that is deficient in the adenoviral function encoded by the cell line. There also have been reports of complementation of replication deficient adenoviral vectors by "helping" vectors, e.g., wild-type virus or conditionally defective mutants.

Replication-deficient adenoviral vectors can be complemented, in trans, by helper virus. This observation alone does not permit isolation of the replication-deficient vectors, however, since the presence of helper virus, needed to provide replicative functions, would contaminate any preparation. Thus, an additional element was needed that would add specificity to the replication and/or packaging of the replication-deficient vector. That element, as provided for in the present invention, derives from the packaging function of adenovirus.

It has been shown that a packaging signal for adenovirus exists in the left end of the conventional adenovirus map (Tibbetts, 1977). Later studies showed that a mutant with a deletion in the EIA (194-358 bp) region of the genome grew poorly even in a cell line that complemented the early (EIA) function (Hearing and Shenk, 1983). When a compensating adenoviral DNA (0-353 bp) was recombined into the right end of the mutant, the virus was packaged normally. Further mutational analysis identified a short, repeated, position-dependent element in the left end of the Ad5 genome. One copy of the repeat was found to be sufficient for efficient packaging if present at either end of the genome, but not when moved towards the interior of the Ad5 DNA molecule (Hearing et al., 1987).

By using mutated versions of the packaging signal, it is possible to create helper viruses that are packaged with varying efficiencies. Typically, the mutations are point mutations or deletions. When helper viruses with low efficiency packaging are grown in helper cells, the virus is packaged, albeit at reduced rates compared to wild-type virus, thereby permitting propagation of the helper. When these helper viruses are grown in cells along with virus that contains wild-type packaging signals, however, the wild-type packaging signals are recognized preferentially over the mutated versions. Given a limiting amount of packaging factor, the virus containing the wild-type signals are packaged selectively when compared to the helpers. If the preference is great enough, stocks approaching homogeneity should be achieved.

Retrovirus

The retroviruses are a group of single-stranded RNA viruses characterized by an ability to convert their RNA to double-stranded DNA in infected cells by a process of

reverse-transcription (Coffin, 1990). The resulting DNA then stably integrates into cellular chromosomes as a provirus and directs synthesis of viral proteins.

The integration results in the retention of the viral gene sequences in the recipient cell and its descendants. The retroviral genome contains three genes - gag, pol and env - that code for capsid proteins, polymerase enzyme, and envelope components, respectively. A sequence found upstream from the gag gene, termed T, functions as a signal for packaging of the genome into virions. Two long terminal repeat (LTR) sequences are present at the 5' and 3' ends of the viral genome. These contain strong promoter and enhancer sequences and also are required for integration in the host cell genome (Coffin, 1990).

In order to construct a retroviral vector, a nucleic acid encoding a promoter is inserted into the viral genome in the place of certain viral sequences to produce a virus that is replication-defective. In order to produce virions, a packaging cell line containing the gag, pol and env genes but without the LTR and T components is constructed (Mann et al., 1983). When a recombinant plasmid containing a human cDNA, together with the retroviral LTR and T sequences is introduced into this cell line (by calcium phosphate precipitation for example), the T sequence allows the RNA transcript of the recombinant plasmid to be packaged into viral particles, which are then secreted into the culture media (Nicolas and Rubenstein, 1988; Temin, 1986; Mann et al., 1983, the disclosures of which are incorporated herein by reference). The media containing the recombinant retroviruses is collected, optionally concentrated, and used for gene transfer. Retroviral vectors are able to infect a broad variety of cell types. However, integration and stable expression of many types of retroviruses require the division of host cells (Paskind et al., 1975).

An approach designed to allow specific targeting of retrovirus vectors recently was developed based on the chemical modification of a retrovirus by the chemical addition of galactose residues to the viral envelope. This modification could permit the specific infection of cells such as hepatocytes via asialoglycoprotein receptors, should this be desired.

A different approach to targeting of recombinant retroviruses was designed in which biotinylated antibodies against a retroviral envelope protein and against a specific cell receptor were used. The antibodies were coupled via the biotin components by

using streptavidin (Roux et al., 1989). Using antibodies against major histocompatibility complex class I and class II antigens, the infection of a variety of human cells that bore those surface antigens was demonstrated with an ecotropic virus in vitro (Roux et al., 1989).

Adeno-associated Virus

AAV utilizes a linear, single-stranded DNA of about 4700 base pairs. Inverted terminal repeats flank the genome. Two genes are present within the genome, giving rise to a number of distinct gene products. The first, the cap gene, produces three different virion proteins (VP), designated VP-1, VP 2 and VP-3.

The second, the rep gene, encodes four non-structural proteins (NS). One or more of these rep gene products is responsible for transactivating AAV transcription.

The three promoters in AAV are designated by their location, in map units, in the genome. These are, from left to right, p5, p19 and p40. Transcription gives rise to six transcripts, two initiated at each of three promoters, with one of each pair being spliced.

The splice site, derived from map units 42-46, is the same for each transcript. The four non-structural proteins apparently are derived from the longer of the transcripts, and three virion proteins all arise from the smallest transcript.

AAV is not associated with any pathologic state in humans. Interestingly, for efficient replication, AAV requires "helping" functions from viruses such as herpes simplex virus I and II, cytornegalovirus, pseudorabies virus and, of course, adenovirus.

The best characterized of the helpers is adenovirus, and many "early" functions for this virus have been shown to assist with AAV replication. Low level expression of AAV rep proteins is believed to hold AAV structural expression in check, and helper virus infection is thought to remove this block.

The terminal repeats of the AAV vector can be obtained by restriction endonuclease digestion of AAV or a plasmid such as p201, which contains a modified AAV genome (Samulski et al, 1987), or by other methods known to the skilled artisan, including but not limited to chemical or enzymatic synthesis of the terminal repeats based upon the published sequence of AAV. The ordinarily skilled artisan can determine, by well-known methods such as deletion analysis, the minimum sequence or

part of the AAV ITRs which is required to allow function, i.e., stable and site specific integration.

The ordinarily skilled artisan also can determine which minor modifications of the sequence can be tolerated while maintaining the ability of the terminal repeats to direct stable, site-specific integration.

AAV-based vectors have proven to be safe and effective vehicles for gene delivery in vitro, and these vectors are being developed and tested in pre-clinical and clinical stages for a wide range of applications in potential gene therapy, both ex vivo and in vivo (Carter and Flotte, 1996; Chattedee et al., 1995; Ferrari et al., 1996; Fisher et al., 1996; Flotte et al., 1993; Goodman et al., 1994; Kaplitt et al., 1994; 1996, Kessler et al., 1996; Koeberl et al., 1997; Mizukami et al., 1996; Xiao et al., 1996, the disclosures of which are incorporated herein by reference in their entireties).

AAV-mediated efficient gene transfer and expression in the lung has led to clinical trials for the treatment of cystic fibrosis (Carter and Flotte, 1996; Flotte et al., 1993, the disclosures of which are incorporated herein by reference). Similarly, the prospects for treatment of muscular dystrophy by AAV-mediated gene delivery of the dystrophin gene to skeletal muscle, of Parkinson's disease by tyrosine hydroxylase gene delivery to the brain, of hemophilia B by Factor IX gene delivery to the liver, and potentially of myocardial infarction by vascular endothelial growth factor gene to the heart, appear promising since AAV-mediated transgene expression in these organs has recently been shown to be highly efficient (Fisher et al., 1996; Flotte et al., 1993; Kaplitt et al., 1994; 1996; Koeberl et al., 1997; McCown et al., 1996; Ping et al., 1996; and Xiao et al., 1996, the disclosures of which are incorporated herein by reference in their entireties.).

Other Viral Vectors

Other viral vectors may be employed as expression constructs in the present invention. Vectors derived from viruses such as vaccinia virus (Ridgeway, 1988; Baichwal and Sugden, 1986; Coupar et al., 1988) and hepatitus B viruses have also been developed and are useful in the present invention. They offer several attractive features for various mammalian cells (Friedmann, 1989; Ridgeway, 1988; Baichwal and

Sugden, 1986; Coupar et al., 1988; and Horwich et al., 1990, the disclosures of which are incorporated herein by reference in their entireties.).

With the recent recognition of defective hepatitis B viruses, new insight was gained into the structure-function relationship of different viral sequences. In vitro studies showed that the virus could retain the ability for helper dependent packaging and reverse transcription despite the deletion of up to 80% of its genome (Horwich et al., 1990). This suggested that large portions of the genome could be replaced with foreign genetic material. Chang et al., recently introduced the chloramphenicol acetyltransferase (CAT) gene into duck hepatitis B virus genome in the place of the polymerase, surface, and pre-surface coding sequences. It was cotransfected with wild-type virus into an avian hepatoma cell line. Culture media containing high titers of the recombinant virus were used to infect primary duckling hepatocytes. Stable CAT gene expression was detected for at least 24 days after transfection (Chang et al., 1991).

In still further embodiments of the present invention, the nucleic acids to be delivered are housed within an infective virus that has been engineered to express a specific binding ligand. The virus particle will thus bind specifically to the cognate receptors of the target cell and deliver the contents to the cell. A novel approach designed to allow specific targeting of retrovirus vectors was recently developed based on the chemical modification of a retrovirus by the chemical addition of lactose residues to the viral envelope. This modification can permit the specific infection of hepatocytes via sialoglycoprotein receptors.

Another approach to targeting of recombinant retroviruses was designed in which biotinylated antibodies against a retroviral envelope protein and against a specific cell receptor were used. The antibodies were coupled via the biotin components by using streptavidin (Roux et al., 1989). Using antibodies against major histocompatibility complex class I and class II antigens, they demonstrated the infection of a variety of human cells that bore those surface antigens with an ecotropic virus in vitro (Roux et al., 1989).

(ii) Non-viral Transfer

DNA constructs of the present invention are generally delivered to a cell. In certain situations, the nucleic acid to be transferred is non-infectious, and can be transferred using non-viral methods.

Several non-viral methods for the transfer of expression constructs into cultured mammalian cells are contemplated by the present invention. These include calcium phosphate precipitation (Graham and Van Der Eb, 1973; Chen and Okayama, 1987; Rippe et al., 1990) DEAE-dextran (Gopal, 1985), electroporation (Tur Kaspa et al., 1986; Potter et al., 1984), direct microinjection (Harland and Weintraub, 1985), DNA loaded liposomes (Nicolau and Sene, 1982; Fraley et al., 1979), cell sonication (Fechheimer et al., 1987), gene bombardment using high velocity microprojectiles (Yang et al., 1990), and receptor-mediated transfection (Wu and Wu, 1987; Wu and Wu, 1988), the disclosures of which are incorporated herein by reference in their entireties.

Once the construct has been delivered into the cell the nucleic acid encoding the therapeutic gene may be positioned and expressed at different sites. In certain embodiments, the nucleic acid encoding the therapeutic gene may be stably integrated into the genome of the cell. Thi–s integration may be in the cognate location and orientation via homologous recombination (gene replacement) or it may be integrated in a random, non-specific location (gene augmentation). In yet further embodiments, the nucleic acid may be stably maintained in the cell as a separate, episomal segment of DNA. Such nucleic acid segments or "episomes" encode sequences sufficient to permit maintenance and replication independent of or in synchronization with the host cell cycle.

How the expression construct is delivered to a cell and where in the cell the nucleic acid remains is dependent on the type of expression construct employed.

In a particular embodiment of the invention, the expression construct may be entrapped in a liposome. Liposomes are vesicular structures characterized by a phospholipid bilayer membrane and an inner aqueous medium. Multilamellar liposomes have multiple lipid layers separated by aqueous medium. They form spontaneously when phospholipids are suspended in an excess of aqueous solution. The lipid components undergo self-rearrangement before the formation of closed structures and

entrap water and dissolved solutes between the lipid bilayers (Ghosh and Bachhawat, 1991). The addition of DNA to cationic liposomes causes a topological transition from liposomes to optically birefringent liquid-crystalline condensed globules (Radler et al., 1997). These DNA-lipid complexes are potential non-viral vectors for use in gene therapy.

Liposome-mediated nucleic acid delivery and expression of foreign DNA in vitro has been very successful. Using the P-lactamase gene, Wong et al. (1980) demonstrated the feasibility of liposome-mediated delivery and expression of foreign DNA in cultured chick embryo, HeLa, and hepatoma cells. Nicolau et al. (1987) accomplished successful liposome-mediated gene transfer in rats after intravenous injection. Also included are various commercial approaches involving "lipofection" technology.

In certain embodiments of the invention, the liposome may be complexed with a hemagglutinating virus (HVJ). This has been shown to facilitate fusion with the cell membrane and promote cell entry of liposome-encapsulated DNA (Kaneda et al., 1989).

In other embodiments, the liposome may be complexed or employed in conjunction with nuclear nonhistone chromosomal proteins (HMG-1) (Kato et al., 1991). In yet further embodiments, the liposome may be complexed or employed in conjunction with both HVJ and HMG-1. In that such expression constructs have been successfully employed in transfer and expression of nucleic acid in vitro and in vivo, then they are applicable for the present invention.

Other vector delivery systems which can be employed to deliver a nucleic acid encoding a therapeutic gene into cells are receptor-mediated delivery vehicles. These take advantage of the selective uptake of macromolecules by receptor mediated endocytosis in almost all eukaryotic cells. Because of the cell type specific distribution of various receptors, the delivery can be highly specific (Wu and Wu, 1993).

Receptor-mediated gene targeting vehicles generally consist of two components: a cell receptor-specific ligand and a DNA-binding agent. Several ligands have been used for receptor-mediated gene transfer. The most extensively characterized ligands are asialoorosomucoid (ASOR) (Wu and Wu, 1987) and transferring (Wagner et al., 1990).

Recently, a synthetic neoglycoprotein, which recognizes the same receptor as ASOR, has been used as a gene delivery vehicle (Ferkol et al., 1993; Perales et al., 1994) and epidermal growth factor (EGF) has also been used to deliver genes to squamous carcinoma cells (Myers, EPO 0273085).

In other embodiments, the delivery vehicle may comprise a ligand and a liposome. For example, Nicolau et al, (1987) employed lactosyl-ceramide, a galactose terminal asialganglioside, incorporated into liposomes and observed an increase in the uptake of the insulin gene by hepatocytes. Thus, it is feasible that a nucleic acid encoding a therapeutic gene also may be specifically delivered into a cell type such as prostate, epithelial or tumor cells, by any number of receptor-ligand systems with or without liposomes. For example, the human prostate-specific antigen (Watt et al, 1986) may be used as the receptor for mediated delivery of a nucleic acid in prostate tissue.

In another embodiment of the invention, the expression construct may simply consist of naked recombinant DNA or plasmids. Transfer of the construct may be performed by any of the methods mentioned above which physically or chemically permeabilize the cell membrane. This is applicable particularly for transfer in vitro, however, it may be applied for in vivo use as well. Dubensky et al, (1984) successfully injected polyornavirus DNA in the form of CaP04 precipitates into liver and spleen of adult and newborn mice demonstrating active viral replication and acute infection.

Benvenisty and Neshif (1986) also demonstrated that direct intraperitoneal injection of CaP04 precipitated plasmids results in expression of the transfected genes. It is envisioned that DNA encoding a CAM may also be transferred in a similar manner in vivo and express CAM.

Another embodiment of the invention for transferring a naked DNA expression construct into cells may involve particle bombardment. This method depends on the ability to accelerate DNA coated microprojectiles to a high velocity allowing them to pierce cell membranes and enter cells without killing them (Klein et al, 1987). Several devices for accelerating small particles have been developed. One such device relies on a high voltage discharge to generate an electrical cur-rent, which in turn provides the motive force (Yang et al, 1990). The microprojectiles used have consisted of biologically inert substances such as tungsten or gold beads.

Antibodies

Polyclonal anti-THAP-family or anti-THAP domain antibodies can be prepared as described above by immunizing a suitable subject with a THAP-family or THAP domain immunogen. The anti-THAP-family or anti- THAP domain antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized THAP-family or THAP domain protein. If desired, the antibody molecules directed against THAP-family can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, e.g., when the anti-THAP-family antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as those described in the following references, the disclosures of which are incorporated herein by reference in their entireties: the hybridoma technique originally described by Kohler and Milstein (1975) Nature 256:495-497) (see also, Brown et al. (1981) J. Immunol. 127:539-46; Brown et al. (1980) J. Biol. Chem. 255:4980-83 ; Yeh et al. (1976) PNAS 76:2927-31; and Yeh et al. (1982) Int. J. Cancer 29:269-75), the more recent human B cell hybridoma technique (Kozbor et al. (1983) Immunol Today 4:72), the EBV-hybridoma technique (Cole et al. (1985), Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally R. H. Kenneth, in Monoclonal Antibodies: A New Dimension In Biological Analyses, Plenum Publishing Corp., New York, N.Y. (1980); E. A. Lerner (1981) Yale J. Biol. Med., 54:387-402; M. L. Gefter et al. (1977) Somatic Cell Genet. 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a THAP-family immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds THAP-family.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-THAP-family or anti-THAP domain monoclonal antibody (see, e.g., G. Galfre et al. (1977) Nature 266:55052; Gefter et al. Somatic Cell Genet., cited supra; Lerner, Yale J Biol.

Med, cited supra; Kenneth, Monoclonal Antibodies, cited supra), the disclosures of which are incorporated herein by reference in their entireties. Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (e.g., a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, e.g., the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind a THAP-family or THAP domain protein, e.g., using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-THAP-family or anti-THAP domain antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with THAP-family or THAP domain protein to thereby isolate immunoglobulin library members that bind THAP-family or THAP domain proteins. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia Recombinant Phage Antibody System, Catalog No. 27-9400-01; and the Stratagene SurfZAP.TM. Phage Display Kit, Catalog No. 240612), the disclosures of which are incorporated herein by reference in their entireties. Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner et al. U.S. Pat. No. 5,223,409; Kang et al. PCT International Publication No. WO 92/18619; Dower et al. PCT International Publication No. WO 91/17271; Winter et al.

PCT International Publication WO 92/20791; Markland et al. PCT International Publication No. WO 92/15679; Breitling et al. PCT International Publication WO 93/01288; McCafferty et al. PCT International Publication No. WO 92/01047; Garrard et al. PCT International Publication No. WO 92/09690; Ladner et al. PCT International Publication No. WO 90/02809; Fuchs et al. (1991) Bio/Technology 9:1370-1372; Hay et al. (1992) Hum. Antibod. Hybridomas 3:81-85; Huse et al. (1989) Science 246:1275-1281; Griffiths et al. (1993) EMBO J 12:725-734; Hawkins et al. (1992) J. Mol. Biol. 226:889-896; Clarkson et al. (1991) Nature 352:624-628; Gram et al. (1992) PNAS 89:3576-3580; Garrad et al. (1991) Bio/Technology 9:1373-1377; Hoogenboom et al. (1991) Nuc. Acid Res. 19:4133-4137; Barbas et al. (1991) PNAS 88:7978-7982; and McCafferty et al. Nature (1990) 348:552-554, the disclosures of which are incorporated herein by reference in their entireties.

Additionally, recombinant anti-THAP-family or anti-THAP domain antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson et al. International Application No. PCT/US86/02269; Akira, et al. European Patent Application 184,187; Taniguchi, M., European Patent Application 171496; Morrison et al. European Patent Application 173,494; Neuberger et al. PCT International Publication No. WO 86/01533; Cabilly et al. U.S. Pat. No. 4,816,567; Cabilly et al. European Patent Application 125,023; Better et al. (1988) Science 240:1041-1043; Liu et al. (1987) PNAS 84:3439-3443; Liu et al. (1987) J. Immunol. 139:3521-3526; Sun et al. (1987) PNAS 84:214-218; Nishimura et al. (1987) Canc. Res. 47:999-1005; Wood et al. (1985) Nature 314:446-449; and Shaw et al. (1988) J. Natl. Cancer Inst. 80:1553-1559); Morrison, S. L. (1985) Science 229:1202-1207; Oi et al. (1986) BioTechniques 4:214; Winter U.S. Pat. No. 5,225,539; Jones et al. (1986) Nature 321:552-525; Verhoeyan et al. (1988) Science 239:1534; and Beidler et al. (1988) J. Immunol. 141:4053-4060, the disclosures of which are incorporated herein by reference in their entireties.

An anti-THAP-family of anti-THAP domain antibody (e.g., monoclonal antibody) can be used to isolate THAP-family or THAP domain protein by standard

techniques, such as affinity chromatography or immunoprecipitation. For example, an anti-THAP-family antibody can facilitate the purification of natural THAP-family from cells and of recombinantly produced THAP-family expressed in host cells. Moreover, an anti-THAP-family antibody can be used to detect THAP-family protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the THAP-family protein. Anti-THAP-family antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

DRUG SCREENING ASSAYS

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g., preferably small molecules, but also peptides, peptidomimetics or other drugs) which bind to THAP-family or THAP domain proteins, have an inhibitory or activating effect on, for example, THAP-family expression or preferably THAP-family activity, or have an inhibitory or activating effect on, for example, the activity of an THAP-family target molecule. In some embodiments small molecules can be generated using combinatorial chemistry or can be obtained from a natural products library. Assays may be cell based, non-cell-based or in vivo assays. Drug screening assays may be binding assays or more preferentially functional assays, as further described.

In general, any suitable activity of a THAP-family protein can be detected in a drug screening assay, including: (1) mediating apoptosis or cell proliferation when expressed or introduced into a cell, most preferably inducing or enhancing apoptosis, and/or most preferably reducing cell proliferation; (2) mediating apoptosis or cell proliferation of an endothelial cell; (3) mediating apoptosis or cell proliferation of a hyperproliferative cell; (4) mediating apoptosis or cell proliferation of a CNS cell, preferably a neuronal or glial cell; (5) an activity indicative of a biological function in an animal selected from the group consisting of mediating, preferably inhibiting angiogenesis, mediating, preferably inhibiting inflammation, inhibition of metastatic potential of cancerous tissue, reduction of tumor burden, increase in sensitivity to chemotherapy or radiotherapy, killing a cancer cell, inhibition of the growth of a cancer cell, or induction of tumor regression; or (6) interaction with a THAP family target molecule or THAP domain target molecule, preferably interaction with a protein or a nucleic acid.

The invention also provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g., preferably small molecules, but also peptides, peptidomimetics or other drugs) which bind to THAP1, PAR4 or PML-NB proteins, and have an inhibitory or activating effect on PAR4 or THAP1 recruitment or binding to or association with PML-NBs or interaction, such as binding, of SLC with a THAP-family polypeptide or a cellular response to SLC which is mediated by a THAP-family polypeptide.

In one embodiment, the invention provides assays for screening candidate or test compounds which are target molecules of a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the `one-bead one-compound` library method; and synthetic library methods using affinity chromatography selection. The biological library

approach is used with peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K. S. (1997) Anticancer Drug Des. 12:145, the disclosure of which is incorporated herein by reference in its entirety).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al. (1993) Proc. Natl. Acad. Sci. U.S.A. 90:6909; Erb et al. (1994) Proc. Natl. Acad. Sci. USA 91:11422; Zuckermann et al. (1994). J. Med. Chem. 37:2678; Cho et al. (1993) Science 261:1303; Carrell et al. (1994) Angew. Chem. Int. Ed. Engl. 33:2059; Carell et al. (1994) Angew. Chem. Int. Ed. Engl. 33:2061; and in Gallop et al. (1994) J. Med. Chem. 37:1233, the disclosures of which are incorporated herein by reference in their entireties.

Libraries of compounds may be presented in solution (e.g., Houghten (1992) Biotechniques 13:412-421), or on beads (Lam (1991) Nature 354:82-84), chips (Fodor (1993) Nature 364:555-556), bacteria (Ladner U.S. Pat. No. 5,223,409), spores (Ladner U.S. Pat. No. '409), plasmids (Cull et al. (1992) Proc Natl Acad Sci USA 89:1865-1869) or on phage (Scott and Smith (1990) Science 249:386-390); (Devin (1990) Science 249:404-406); (Cwirla et al. (1990) Proc. Natl. Acad. Sci. 87:6378-6382); (Felici (1991) J. Mol. Biol. 222:301-310); (Ladner supra.), the disclosures of which are incorporated herein by reference in their entireties.

Determining the ability of the test compound to inhibit or increase THAP-family polypeptide activity can also be accomplished, for example, by coupling the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof with a radioisotope or enzymatic label such that binding of the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof to its cognate target molecule can be determined by detecting the labeled THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof in a complex. For example, compounds (e.g., THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof) can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an

appropriate substrate to product. The labeled molecule is placed in contact with its cognate molecule and the extent of complex formation is measured. For example, the extent of complex formation may be measured by immuno precipitating the complex or by performing gel electrophoresis.

It is also within the scope of this invention to determine the ability of a compound (e.g., THAP family or THAP domain polypeptide, or biologically active fragment or homologue thereof) to interact with its cognate target molecule without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a compound with its cognate target molecule without the labeling of either the compound or the target molecule. McConnell, H. M. et al. (1992) Science 257:1906-1912, the disclosure of which is incorporated herein by reference in its entirety. A microphysiometer such as a cytosensor is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between compound and cognate target molecule.

In a preferred embodiment, the assay comprises contacting a cell which expresses a THAP family or THAP domain polypeptide, or biologically active fragment or homologue thereof, with a THAP-family or THAP domain protein target molecule to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to inhibit or increase the activity of the THAP family or THAP domain polypeptide, or biologically active fragment or homologue thereof, wherein determining the ability of the test compound to inhibit or increase the activity of the THAP family or THAP domain polypeptide, or biologically active fragment or homologue thereof, comprises determining the ability of the test compound to inhibit or increase a biological activity of the THAP-family polypeptide expressing cell.

In another embodiment, the assay comprises contacting a cell which expresses a THAP family or THAP domain polypeptide, or biologically active fragment or homologue thereof, with a test compound, and determining the ability of the test compound to inhibit or increase the activity of the THAP family or THAP domain polypeptide, or biologically active fragment or homologue thereof, wherein determining the ability of the test compound to inhibit or increase the activity of the THAP family or

THAP domain polypeptide, or biologically active fragment or homologue thereof, comprises determining the ability of the test compound to inhibit or increase a biological activity of the THAP-family polypeptide expressing cell.

In another preferred embodiment, the assay comprises contacting a cell which is responsive to a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof, with a THAP-family protein or biologically-active portion thereof, to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to modulate the activity of the THAP-family protein or biologically active portion thereof, wherein determining the ability of the test compound to modulate the activity of the THAP-family protein or biologically active portion thereof comprises determining the ability of the test compound to modulate a biological activity of the THAP-family polypeptide-responsive cell (e.g., determining the ability of the test compound to modulate a THAP-family polypeptide activity.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a THAP-family target molecule (i.e. a molecule with which THAP-family polypeptide interacts) with a test compound and determining the ability of the test compound to modulate (e.g. stimulate or inhibit) the activity of the THAP-family target molecule. Determining the ability of the test compound to modulate the activity of a THAP-family target molecule can be accomplished, for example, by determining the ability of the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof to bind to or interact with the THAP-family target molecule.

Determining the ability of the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof to bind to or interact with a THAP-family target molecule can be accomplished by one of the methods described above for determining direct binding. In a preferred embodiment, determining the ability of the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof to bind to or interact with a THAP-family target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by contacting the target molecule with the THAP family or THAP domain polypeptide, or a biologically active fragment or

homologue thereof and measuring induction of a cellular second messenger of the target (i.e. intracellular Ca^2+, diacylglycerol, IP_3, etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a target-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, e.g., luciferase), or detecting a target-regulated cellular response, for example, signal transduction or protein:protein interactions.

In yet another embodiment, an assay of the present invention is a cell-free assay in which a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof is contacted with a test compound and the ability of the test compound to bind to the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof is determined. Binding of the test compound to the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof with a known compound which binds THAP-family polypeptide (e.g., a THAP-family target molecule) to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof, wherein determining the ability of the test compound to interact with a THAP-family protein comprises determining the ability of the test compound to preferentially bind to THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof as compared to the known compound.

In another embodiment, the assay is a cell-free assay in which a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof is contacted with a test compound and the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof is determined. Determining the ability of the test compound to modulate the activity of a THAP-family protein can be accomplished, for example, by determining the ability of the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof to bind to a THAP-family target molecule by one of the methods described above for determining

direct binding. Determining the ability of the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof to bind to a THAP-family target molecule can also be accomplished using a technology such as real-time Biomolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) Anal. Chem. 63:2338-2345 and Szabo et al. (1995) Curr. Opin. Struct. Biol. 5:699-705, the disclosures of which are incorporated herein by reference in their entireties. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the optical phenomenon of surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

In an alternative embodiment, determining the ability of the test compound to modulate the activity of a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof can be accomplished by determining the ability of the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof to further modulate the activity of a downstream effector (e.g., a growth factor mediated signal transduction pathway component) of a THAP-family target molecule. For example, the activity of the effector molecule on an appropriate target can be determined or the binding of the effector to an appropriate target can be determined as previously described.

In yet another embodiment, the cell-free assay involves contacting a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof with a known compound which binds the THAP-family protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the THAP-family protein, wherein determining the ability of the test compound to interact with the THAP-family protein comprises determining the ability of the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof to preferentially bind to or modulate the activity of a THAP-family target molecule.

The cell-free assays of the present invention are amenable to use of both soluble and/or membrane-bound forms of isolated proteins (e.g. THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof or molecules to which THAP-family targets bind). In the case of cell-free assays in which a membrane-

bound form an isolated protein is used it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the isolated protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton.[RTM. X-100, Triton.RTM. X-114, Thesit.RTM.], Isotridecypoly(ethylene glycol ether)$_n$,3-[(3-cholamidopropyl)dimethylamminio]- 1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof or a target molecule thereof to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof, or interaction of a THAP-family protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/THAP-family fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, Mo.) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or THAP-family protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of THAP-family polypeptide binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either a THAP-family protein or a THAP-family target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated THAP-family protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with a THAP-family protein or target molecule but which do not interfere with binding of the THAP-family protein to its target molecule can be derivatized to the wells of the plate, and unbound target or THAP-family protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the THAP-family protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the THAP-family protein or target molecule.

In another embodiment, modulators of THAP-family or THAP domain polypeptides expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of THAP-family or THAP domain polypeptides mRNA or protein in the cell is determined. The level of expression of THAP-family polypeptide mRNA or protein in the presence of the candidate compound is compared to the level of expression of THAP-family polypeptide or THAP domain mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of THAP-family polypeptide expression based on this comparison. For example, when expression of THAP-family polypeptide or THAP domain mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of THAP-family polypeptide or THAP domain mRNA or protein expression. Alternatively, when expression of THAP-family polypeptide or THAP domain mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of THAP-family polypeptide or THAP domain mRNA or protein expression. The level of THAP-family polypeptide or THAP domain mRNA or protein expression in the cells

can be determined by methods described herein for detecting THAP-family polypeptide or THAP domain mRNA or protein.

In yet another aspect of the invention, the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay using the methods described above for use in THAP-family polypeptide/PAR4 interactions assays, to identify other proteins which bind to or interact with THAP-family polypeptide ("THAP-family-binding proteins" or "THAP-family-bp") and are involved in THAP-family polypeptide activity. Such THAP-family- or THAP domain-binding proteins are also likely to be involved in the propagation of signals by the THAP-family or THAP domain proteins or THAP-family or THAP domain proteins targets as, for example, downstream elements of a THAP-family polypeptide- or THAP domain-mediated signaling pathway. Alternatively, such THAP-family-binding proteins are likely to be THAP-family polypeptides inhibitors.

THAP/DNA BINDING ASSAYS

In another embodiment of the invention a method is provided for identifying compounds which interfere with THAP-family DNA binding activity, comprising the steps of: contacting a THAP-family protein or a portion thereof immobilized on a solid support with both a test compound and DNA fragments, or contacting a DNA fragment immobilized on a solid support with both a test compound and a THAP-family protein. The binding between DNA and the THAP-protein or a portion thereof is detected, wherein a decrease in DNA binding when compared to DNA binding in the absence of the test compound indicates that the test compound is an inhibitor of THAP-family DNA binding activity, and an increase in DNA binding when compared to DNA binding in the absence of the test compound indicates that the test compound is an inducer of or restores THAP-family DNA binding activity. As discussed further, DNA fragments may be selected to be specific THAP-family protein target DNA obtained for example as described in Example 28, or may be non-specific THAP-family target DNA. Methods for detecting protein-DNA interactios are well known in the art, including most commonly used electrophoretic mobility shift assays (EMSAs) or by filter binding (Zabel et al, (1991) J. Biol. Chem., 266:252; and Okamoto and Beach, (1994) Embo J.

13: 4816). Other assays are available which are amenable for high throughput detection and quantification of specific and nonspecific DNA binding (Amersham, N.J.; and Gal S. et al, 6th Ann. Conf. Soc. Biomol. Screening, 6-9 Sept 2000, Vancouver, B.C.).

In a first aspect, a screening assay involves identifying compounds which interfere with THAP-family DNA binding activity without prior knowledge about specific THAP-family binding sequences. For example, a THAP-family protein is contacted with both a test compound and a library of oligonucleotides or a sample of DNA fragments not selected based on specific DNA sequences. Preferably the THAP-family protein is immobilized on a solid support (such as an array or a column). Unbound DNA is separated from DNA which is bound to the THAP-famliy protein, and the DNA which is bound to THAP-family protein is detected and can be quantitated by any means known in the art. For example, the DNA fragment is labelled with a detectable moiety, such as a radioactive moiety, a colorimetric moiety or a fluorescent moiety. Techniques for so labelling DNA are well known in the art.

The DNA which is bound to the THAP-family protein or a portion thereof is separated from unbound DNA by immunoprecipitation with antibodies which are specific for the THAP-family protein or a portion thereof. Use of two different monoclonal anti-THAP-family antibodies may result in more complete immunoprecipitation than either one alone. The amount of DNA which is in the immunoprecipitate can be quantitated by any means known in the art. THAP-family proteins or portions thereof which bind to the DNA can also be detected by gel shift assays (Tan, Cell, 62:367, 1990), nuclease protection assays, or methylase interference assays.

It is still another object of the invention to provide methods for identifying compounds which restore the ability of mutant THAP-family proteins or portions thereof to bind to DNA sequences. In one embodiment a method of screening agents for use in therapy is provided comprising: measuring the amount of binding of a THAP-family protein or a portion thereof which is encoded by a mutant gene found in cells of a patient to DNA molecules, preferably random oligonucleotides or DNA fragments from a nucleic acid library; measuring the amount of binding of said THAP-family protein or a portion thereof to said nucleic acid molecules in the presence of a test substance; and comparing the amount of binding of the THAP-family protein or a

portion thereof in the presence of said test substance to the amount of binding of the THAP-family protein in the absence of said test substance, a test substance which increases the amount of binding being a candidate for use in therapy.

In another embodiment of the invention, oligonucleotides can be isolated which restore to mutant THAP-family proteins or portions thereof the ability to bind to a consensus binding sequence or conforming sequences. Mutant THAP-family protein or a portion thereof and random oligonucleotides are added to a solid support on which THAP-family-specific DNA fragments are immobilized. Oligonucleotides which bind to the solid support are recovered and analyzed. Those whose binding to the solid support is dependent on the presence of the mutant THAP-family protein are presumptively binding the support by binding to and restoring the conformation of the mutant protein.

If desired, specific binding can be distinguished from non-specific binding by any means known in the art. For example, specific binding interactions are stronger than non-specific binding interactions. Thus the incubation mixture can be subjected to any agent or condition which destabilizes protein/DNA interactions such that the specific binding reaction is the predominant one detected. Alternatively, as taught more specifically below, a non-specific competitor, such as dI-dC, can be added to the incubation mixture. If the DNA containing the specific binding sites is labelled and the competitor is unlabeled, then the specific binding reactions will be the ones predominantly detected upon measuring labelled DNA.

According to another embodiment of the invention, after incubation of THAP-family protein or a portion thereof with specific DNA fragments all components of the cell lysate which do not bind to the DNA fragments are removed. This can be accomplished, among other ways, by employing DNA fragments which are attached to an insoluble polymeric support such as agarose, cellulose and the like. After binding, all non-binding components can be washed away, leaving THAP-family protein or a portion thereof bound to the DNA/solid support. The THAP-family protein or a portion thereof can be quantitated by any means known in the art. It can be determined using an immunological assay, such as an ELISA, RIA or Western blotting.

In another embodiment of the invention a method is provided for identifying compounds which specifically bind to THAP-family-specific-DNA sequences,

comprising the steps of: contacting a THAP-family-specific DNA fragment immobilized on a solid support with both a test compound and wild-type THAP-family protein or a portion thereof to bind the wild-type THAP-family protein or a portion thereof to the DNA fragment; determining the amount of wild-type THAP-family protein which is bound to the DNA fragment, inhibition of binding of wild-type THAP-family protein by the test compound with respect to a control lacking the test compound suggesting binding of the test compound to the THAP-family-specific DNA binding sequences.

It is still another object of the invention to provide methods for identifying compounds which restore the ability of mutant THAP-family proteins or portions thereof to bind to specific DNA binding sequences. In one embodiment a method of screening agents for use in therapy is provided comprising: measuring the amount of binding of a THAP-family protein or a portion thereof which is encoded by a mutant gene found in cells of a patient to a DNA molecule which comprises more than one monomer of a specific THAP-family target nucleotide sequence; measuring the amount of binding of said THAP-family protein to said nucleic acid molecule in the presence of a test substance; and comparing the amount of binding of the THAP-family protein in the presence of said test substance to the amount of binding of the THAP-family protein or a portion thereof in the absence of said test substance, a test substance which increases the amount of binding being a candidate for use in therapy.

In another embodiment of the invention a method is provided for screening agents for use in therapy comprising: contacting a transfected cell with a test substance, said transfected cell containing a THAP-family protein or a portion thereof which is encoded by a mutant gene found in cells of a patient and a reporter gene construct comprising a reporter gene which encodes an assayable product and a sequence which conforms to a THAP-family DNA binding site, wherein said sequence is upstream from and adjacent to said reporter gene; and determining whether the amount of expression of said reporter gene is altered by the test substance, a test substance which alters the amount of expression of said reporter gene being a candidate for use in therapy.

In still another embodiment a method of screening agents for use in therapy is provided comprising: adding RNA polymerase ribonucleotides and a THAP-family protein or a portion thereof to a transcription construct, said transcription construct

comprising a reporter gene which encodes an assayable product and a sequence which conforms to a THAP-family consensus binding site, said sequence being upstream from and adjacent to said reporter gene, said step of adding being effected in the presence and absence of a test substance; determining whether the amount of transcription of said reporter gene is altered by the presence of said test substance, a test substance which alters the amount of transcription of said reporter gene being a candidate for use in therapy.

According to the present invention compounds which have THAP-family activity are those which specifically complex with a THAP-family-specific DNA binding site. Oligonucleotides and oligonucleotide containing nucleotide analogs are also contemplated among those compounds which are able to complex with a THAP-family-specific DNA binding site.

Further assays to modulate THAP-family polypeptide activity in vivo

It will be appreciated that any suitable assay that allows detection of THAP-family polypeptide or THAP domain activity can be used. Examples of assays for testing protein interaction, nucleic acid binding or modulation of apoptosis in the presence or absence of a test compound are further described herein. Thus, the invention encompasses a method of identifying a candidate THAP-family polypeptide modulator (e.g. activator or inhibitor), said method comprising:

a) providing a cell comprising a THAP family or THAP domain polypeptide, or a biologically active fragment or homolog thereof;

b) contacting said cell with a test compound; and

c) determining whether said compound selectively modulates (e.g. activates or inhibits) THAP-family polypeptide activity, preferably pro-apoptotic activity, or THAP family or THAP domain target binding; wherein a determination that said compound selectively modulates (e.g. activates or inhibits) the activity of said polypeptide indicates that said compound is a candidate modulator (e.g. activator or inhibitor respectively) of said polypeptide. Preferably, the THAP family or THAP domain target is a protein or nucleic acid.

Preferably the cell is a cell which has been transfected with an recombinant expression vector encoding a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof.

Several examples of assays for the detection of apoptosis are described herein, in the section titled "Apoptosis assays". Several examples of assays for the detection of THAP family or THAP domain target interactions are described herein, including assays for detection of protein interactions and nucleic acid binding.

In one example of an assay for apoptosis activity, a high throughput screening assay for molecules that abrogate or stimulate THAP-family polypeptide proapoptotic activity is provided based on serum-withdrawal induced apoptosis in a 3T3 cell line with tetracycline-regulated expression of a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof. Apoptotic cells can be detected by TUNEL labeling in 96- or 384-wells microplates. A drug screening assay can be carried out along the lines as described in Example 23. 3T3 cells, which have previously been used to analyze the pro-apoptotic activity of PAR4 (Diaz-Meco et al, 1996; Berra et al., 1997), can be transfected with expression vectors encoding a THAP-family or THAP domain polypeptide allowing the ectopic expression of THAP-family polypeptide. Then, the apoptotic response to serum withdrawal is assayed in the presence of a test compound, allowing the identification of test compounds that either enhance or inhibit the ability of THAP-family or THAP domain polypeptide to induce apoptosis. Transfected cells are deprived of serum and cells with apoptotic nuclei are counted. Apoptotic nuclei can be counted by DAPI staining and in situ TUNEL assays.

Further THAP-family polypeptide/THAP-target interaction assays

In exemplary methods THAP/THAP target interaction assays are described in the context of THAP1 and the THAP target Par4. However, it will be appreciated that assays for screening for modulators of other THAP family members or THAP domains and other THAP target molecules may be carried out by substituting these for THAP1 and Par4 in the methods below. For example, in some embodiments, modulators which affect the interaction between a THAP-family polypeptide and SLC are identified.

As demonstrated in Examples 4, 5, 6, and 7 and Figures 3, 4 and 5, the inventors have demonstrated using several experimental methods that THAP1 interacts with the

pro-apoptotic protein Par4. In particular, it has been shown that THAP1 interacts with Par4 wild type (Par4) and a Par4 death domain (Par4DD) in a yeast two-hybrid system. Yeast cells were cotransformed with BD7-THAP1 and AD7-Par4, AD7, AD7-Par4DD or AD7-Par4) expression vectors. Transformants were selected on media lacking histidine and adenine. Identical results were obtained by cotransformation of AD7-THAP1 with BD7-Par4, BD7, BD7-Par4DD or BD7-Par4).

The inventors have also demonstrated *in vitro* binding of THAP1 to GST-Par4DD. Par4DD was expressed as a GST fusion protein, purified on glutathione sepharose and employed as an affinity matrix for binding of *in vitro* translated ^{35}S-methionine labeled THAP1. GST served as negative control.

Futhermore, the inventors have shown that THAP1 interacts with both Par4DD and SLC *in vivo*. Myc-Par4DD and GFP-THAP1 expression vectors were cotransfected in primary human endothelial cells. Myc-Par4DD was stained with monoclonal anti-myc antibody. Green fluorescence, GFP-THAP1; red fluorescence, Par4DD.

The invention thus encompasses assays for the identification of molecules that modulate (stimulate or inhibit) THAP-family polypeptide/PAR4 binding. In preferred embodiments, the invention includes assays for the identification of molecules that modulate (stimulate or inhibit) THAP1 /PAR4 binding or THAP1/SLC binding.

Four examples of high throughput screening assays include:

1) a two hybrid-based assay in yeast to find drugs that disrupt interaction of the THAP-family bait with the PAR4 or SLC as prey

2) an in vitro interaction assay using recombinant THAP-family polypeptide and PAR4 or SLC proteins

3) a chip-based binding assay using recombinant THAP-family polypeptide and PAR4 or SLC proteins

2) a fluorescence resonance energy transfer (FRET) cell-based assay using THAP-family polypeptide and PAR4 or SLC proteins fused with fluorescent proteins

The invention thus encompasses a method of identifying a candidate THAP-family polypeptide/PAR4 or SLC interaction modulator, said method comprising:

a) providing a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof and a PAR4 or SLC polypeptide or fragment thereof;

b) contacting said THAP family or THAP domain polypeptide with a test compound; and

c) determining whether said compound selectively modulates (e.g. activates or inhibits) THAP-family/PAR4 or SLC interaction activity.

Also envisioned is a method comprising:

a) providing a cell comprising a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof and a PAR4 or SLC polypeptide or fragment thereof;

b) contacting said cell with a test compound; and

c) determining whether said compound selectively modulates (e.g. activates or inhibits) THAP-family/PAR4 or SLC interaction activity.

In general, any suitable assay for the detection of protein-protein interaction may be used.

In one example, a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof can be used as a "bait protein" and a PAR4 or SLC protein can be used as a "prey protein" (or vice-versa) in a two-hybrid assay (see, e.g., U.S. Pat. No. 5,283,317; Zervos et al. (1993) Cell 72:223-232; Madura et al. (1993) J. Biol. Chem. 268:12046-12054; Bartel et al. (1993) Biotechniques 14:920-924; Iwabuchi et al. (1993) Oncogene 8:1693-1696; and Brent WO94/10300, the disclosures of which are incorporated herein by reference in their entireties). The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, the gene that codes for a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, in vivo, forming a THAP-family polypeptide/PAR4 complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a

transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the THAP-family protein. This assay can thus be carried out in the presence or absence of a test compound, whereby modulation of THAP-family polypeptide/PAR4 or SLC interaction can be detected by lower or lack of transcription of the reported gene.

In other examples, in vitro THAP-family polypeptide/PAR4 or SLC interaction assays can be carried out, several examples of which are further described herein. For example, a recombinant THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof is contacted with a recombinant PAR4 or SLC protein or biologically active portion thereof, and the ability of the PAR4 or SLC protein to bind to the THAP-family protein is determined. Binding of the PAR4 or SLC protein compound to the THAP-family protein can be determined either directly or indirectly as described herein. In a preferred embodiment, the assay includes contacting the THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof with a PAR4 or SLC protein which binds a THAP-family protein (e.g., a THAP-family target molecule) to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a THAP-family protein, wherein determining the ability of the test compound to interact with a THAP-family protein comprises determining the ability of the test compound to preferentially bind to THAP-family or biologically active portion thereof as compared to the PAR4 or SLC protein. For example, the step of determining the ability of the test compound to interact with a THAP-family protein may comprise determining the ability of the compound to displace Par4 or SLC from a THAP-family protein/Par4 or SLC complex thereby forming a THAP-family protein/compound complex. Alternatively, it will be appreciated that it is also possible to determine the ability of the test compound to interact with a PAR4 or SLC protein, wherein determining the ability of the test compound to interact with a PAR4 or SLC protein comprises determining the ability of the test compound to preferentially bind to PAR4 or SLC or biologically active portion thereof as compared to the THAP-family protein. For example, the step of determining the ability of the test compound to interact with a

THAP-family protein may comprise determining the ability of the compound to displace Par4 or SLC from a THAP-family protein/Par4 or SLC complex thereby forming a THAP-family protein/compound complex.

Assays to modulate THAP-family polypeptide and/or Par4 trafficking in the PML nuclear bodies (PML NBs)

As demonstrated in Examples 8 and 9, the inventors have demonstrated using several experimental methods that THAP1 and Par4 localize in PML NBs.

The inventors demonstrated that THAP1 is a novel protein associated with PML-nuclear bodies. Double immunofluorescence staining showed colocalization of THAP1 with PML-NBs proteins, PML and Daxx. Primary human endothelial cells were transfected with GFP-THAP1 expression vector ; endogenous PML and Daxx were stained with monoclonal anti-PML and polyclonal anti-Daxx antibodies, respectively.

The inventors also demonstrated that Par4 is a novel component of PML-NBs that colocalizes with THAP1 *in vivo* by several experiments. In one experiments, double immunofluorescence staining revealed colocalization of Par4 and PML at PML-NBs in primary human endothelial cells or fibroblasts. Endogenous PAR4 and PML were stained with polyclonal anti-PAR4 and monoclonal anti-PML antibodies, respectively. In another experiment, double staining revealed colocalization of Par4 and THAP1 in cells expressing ectopic GFP-THAP1. Primary human endothelial cells or fibroblasts were transfected with GFP-THAP1 expression vector ; endogenous Par4 was stained with polyclonal anti-PAR4 antibodies.

The inventors further demonstrated that PML recruits the THAP1/Par4 complex to PML-NBs. Triple immunofluorescence staining showed colocalization of THAP1, Par4 and PML in cells overexpressing PML and absence of colocalization in cells expressing ectopic Sp100. Hela cells were cotransfected with GFP-THAP1 and HA-PML or HA-SP100 expression vectors; HA-PML or HA-SP100 and endogenous Par4 were stained with monoclonal anti-HA and polyclonal anti-Par4 antibodies, respectively.

Assays to modulate THAP family protein trafficking in the PML nuclear bodies

Provided are assays for the identification of drugs that modulate (stimulate or inhibit) THAP-family or THAP domain protein, particularly THAP1, binding to PML-NB proteins or localization to PML-NBs. In general, any suitable assay for the detection of protein-protein interaction may be used. Two examples of high throughput screening assays include 1) a two hybrid-based assay in yeast to find compounds that disrupt interaction of the THAP1 bait with the PML-NB protein prey; and 2) in vitro interaction assays using recombinant THAP1 and PML-NB proteins. Such assays may be conducted as described above with respect to THAP-family/Par4 assays except that the PML-NB protein is used in place of Par4. Binding may be detected, for example, between a THAP-family protein and a PML protein or PML associated protein such as daxx, sp100, sp140, p53, pRB, CBP, BLM or SUMO-1.

Other assays for which standard methods are well known include assays to identify molecules that modulate, generally inhibit, the colocalization of THAP1 with PML-NBs. Detection can be carried out using a suitable label, such as an anti-THAP1 antibody, and an antibody allowing the detection of PML-NB protein.

Assays to modulate PAR4 trafficking in the PML bodies

Provided are assays for the identification of drugs that modulate (stimulate or inhibit) PAR4 binding to PML-NB proteins or localization to PML-NBs. In general, any suitable assay for the detection of protein-protein interaction may be used. Two examples of high throughput screening assays include 1) a two hybrid-based assay in yeast to find compounds that disrupt interaction of the PAR4 bait with the PML-NB protein prey; and 2) in vitro interaction assays using recombinant PAR4 and PML-NB proteins. Such assays may be conducted as described above with respect to THAP-family polypeptide/Par4 assays except that the PML-NB protein is used in place of the THAP-family polypeptide. Binding may be detected, for example, between a Par4 protein and a PML protein or PML associated protein such as daxx, sp100, sp140, p53, pRB, CBP, BLM or SUMO-1.

Other assays for which standard methods are well known include assays to identify molecules that modulate, generally inhibit, the colocalization of PAR4 with

PML-NBs. Detection can be carried out using a suitable label, such as an anti-PAR4 antibody, and an antibody allowing the detection of PML-NB protein.

This invention further pertains to novel agents identified by the above-described screening assays and to processes for producing such agents by use of these assays. Accordingly, in one embodiment, the present invention includes a compound or agent obtainable by a method comprising the steps of any one of the aforementioned screening assays (e.g., cell-based assays or cell-free assays). For example, in one embodiment, the invention includes a compound or agent obtainable by a method comprising contacting a cell which expresses a THAP-family target molecule with a test compound and determining the ability of the test compound to bind to, or modulate the activity of, the THAP-family target molecule. In another embodiment, the invention includes a compound or agent obtainable by a method comprising contacting a cell which expresses a THAP-family target molecule with a THAP-family protein or biologically-active portion thereof, to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with, or modulate the activity of, the THAP-family target molecule. In another embodiment, the invention includes a compound or agent obtainable by a method comprising contacting a THAP-family protein or biologically active portion thereof with a test compound and determining the ability of the test compound to bind to, or modulate (e.g., stimulate or inhibit) the activity of, the THAP-family protein or biologically active portion thereof. In yet another embodiment, the present invention includes a compound or agent obtainable by a method comprising contacting a THAP-family protein or biologically active portion thereof with a known compound which binds the THAP-family protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with, or modulate the activity of the THAP-family protein.

Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a THAP-family or THAP domain modulating agent, an antisense THAP-family or THAP domain nucleic acid molecule, a THAP-family- or THAP domain- specific antibody, or a THAP-family- or THAP domain- binding partner) can be used in an animal model to determine the efficacy, toxicity, or side

effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

The present invention also pertains to uses of novel agents identified by the above-described screening assays for diagnoses, prognoses, and treatments as described herein. Accordingly, it is within the scope of the present invention to use such agents in the design, formulation, synthesis, manufacture, and/or production of a drug or pharmaceutical composition for use in diagnosis, prognosis, or treatment, as described herein. For example, in one embodiment, the present invention includes a method of synthesizing or producing a drug or pharmaceutical composition by reference to the structure and/or properties of a compound obtainable by one of the above-described screening assays. For example, a drug or pharmaceutical composition can be synthesized based on the structure and/or properties of a compound obtained by a method in which a cell which expresses a THAP-family target molecule is contacted with a test compound and the ability of the test compound to bind to, or modulate the activity of, the THAP-family target molecule is determined. In another exemplary embodiment, the present invention includes a method of synthesizing or producing a drug or pharmaceutical composition based on the structure and/or properties of a compound obtainable by a method in which a THAP-family protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to, or modulate (e.g., stimulate or inhibit) the activity of, the THAP-family protein or biologically active portion thereof is determined.

Apoptosis assays

It will be appreciated that any suitable apoptosis assay may be used to assess the apoptotic activity of a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof.

Apoptosis can be recognized by a characteristic pattern of morphological, biochemical, and molecular changes. Cells going through apoptosis appear shrunken, and rounded; they also can be observed to become detached from culture dish. The morphological changes involve a characteristic pattern of condensation of chromatin

and cytoplasm which can be readily identified by microscopy. When stained with a DNA-binding dye, e.g., H33258, apoptotic cells display classic condensed and punctate nuclei instead of homogeneous and round nuclei.

A hallmark of apoptosis is endonucleolysis, a molecular change in which nuclear DNA is initially degraded at the linker sections of nucleosomes to give rise to fragments equivalent to single and multiple nucleosomes. When these DNA fragments are subjected to gel electrophoresis, they reveal a series of DNA bands which are positioned approximately equally distant from each other on the gel. The size difference between the two bands next to each other is about the length of one nucleosome, i.e., 120 base pairs. This characteristic display of the DNA bands is called a DNA ladder and it indicates apoptosis of the cell. Apoptotic cells can be identified by flow cytometric methods based on measurement of cellular DNA content, increased sensitivity of DNA to denaturation, or altered light scattering properties. These methods are well known in the art and are within the contemplation of the invention.

Abnormal DNA breaks which are characteristic of apoptosis can be detected by any means known in the art. In one preferred embodiment, DNA breaks are labeled with biotinylated dUTP (b-dUTP). As described in U.S. Patent No. 5,897,999, the disclosure of which is incorporated herein by reference, cells are fixed and incubated in the presence of biotinylated dUTP with either exogenous terminal transferase (terminal DNA transferase assay; TdT assay) or DNA polymerase (nick translation assay; NT assay). The biotinylated dUTP is incorporated into the chromosome at the places where abnormal DNA breaks are repaired, and are detected with fluorescein conjugated to avidin under fluorescence microscopy.

Assessing THAP-family, THAP domain and PAR4 polypeptides activity

For assessing the nucleic acids and polypeptides of the invention, the apoptosis indicator which is assessed in the screening method of the invention may be substantially any indicator of the viability of the cell. By way of example, the viability indicator may be selected from the group consisting of cell number, cell refractility, cell fragility, cell size, number of cellular vacuoles, a stain which distinguishes live cells from dead cells, methylene blue staining, bud size, bud location, nuclear morphology, and nuclear staining. Other viability indicators and combinations of the viability

indicators described herein are known in the art and may be used in the screening method of the invention.

Cell death status can be evaluated based on DNA integrity. Assays for this determination include assaying DNA on an agarose gel to identify DNA breaking into oligonucleosome ladders and immunohistochemically detecting the nicked ends of DNA by labeling the free DNA end with fluorescein or horseradish peroxidase-conjugated UTP via terminal transferase. Routinely, one can also examine nuclear morphology by propidium iodide (PI) staining. All three assays (DNA ladder, end-labeling, and PI labelling) are gross measurements and good for those cells that are already dead or at the end stage of dying.

In a preferred example, an apoptosis assay is based on serum-withdrawal induced apoptosis in a 3T3 cell line with tetracycline-regulated expression of a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof. Detection of apoptotic cells is accomplished by TUNEL labeling cells in 96- or 384-well microplates. This example is further described in Example 23.

In other aspects, assays may test for the generation of cytotoxic death signals, anti-viral responses (Tartaglia et al., (1993) Cell 74(5):845-531), and/or the activation of acid sphingomyelinase (Wiegmann et al., (1994) Cell 78(6):1005-15) when the THAP-family protein is overexpressed or ectopically expressed in cells. Assaying for modulation of apoptosis can also be carried out in neuronal cells and lymphocytes for example, where factor withdrawal is known to induce cell suicide as demonstrated with neuronal cells requiring nerve growth factor to survive (Martin, D. P. et al, (1988) J. Cell Biol 106, 829-844) and lymphocytes depending on a specific lymphokine to live (Kyprianou, N. and Isaacs, J. T. (1988) Endrocrinology 122:552-562). The above disclosures are incorporated herein by reference.

THAP-family or THAP domain polypeptide –marker fusions in cell assays

In one method, an expression vector encoding the a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof can be used to evaluate the ability of the polypeptides of the invention to induce apoptosis in cells. If desired, a THAP-family or THAP domain polypeptide may be fused to a detectable marker in order to facilitate identification of those cells expressing the a THAP family

or THAP domain polypeptide, or a biologically active fragment or homologue thereof. For example, a variant of the Aequoria victoria GFP variant, enhanced green fluorescent protein (EGFP), can be used in fusion protein production (CLONTECH Laboratories, Inc., 1020 East Meadow Circle, Palo Alto, Calif. 94303), further described in U.S. Patent No. 6,191,269, the disclosure of which is incorporated herein by reference.

The THAP-family- or THAP domain polypeptide cDNA sequence is fused in-frame by insertion of the THAP-family- or THAP domain polypeptide encoding cDNA into the SalI-BamHI site of plasmid pEGFP-NI (GenBank Accession # U55762). Cells are transiently transfected by the method optimal for the cell being tested (either $CaPO^4$ or Lipofectin). Expression of a THAP-family or THAP domain polypeptide and induction of apoptosis is examined using a fluorescence microscope at 24 hrs and 48 hrs post-transfection. Apoptosis can be evaluated by the TUNEL method (which involves 3' end-labeling of cleaved nuclear and/or morphological criteria DNA) (Cohen et al. (1984) J. Immunol. 132:38-42, the disclosure of which is incorporated herein by reference). Where the screen uses a fusion polypeptide comprising a THAP-family or THAP domain polypeptide and a reporter polypeptide (e.g., EGFP), apoptosis can be evaluated by detection of nuclear localization of the reporter polypeptide in fragmented nuclear bodies or apoptotic bodies. For example, where a THAP-family or THAP domain polypeptide- EGFP fusion polypeptide is used, distribution of THAP-family or THAP domain polypeptide EGFP-associated fluorescence in apoptotic cells would be identical to the distribution of DAPI or Hoechst 33342 dyes, which are conventionally used to detect the nuclear DNA changes associated with apoptosis (Cohen et al., supra). A minimum of approximately 100 cells, which display characteristic EGFP fluorescence, are evaluated by fluorescence microscopy. Apoptosis is scored as nuclear fragmentation, marked apoptotic bodies, and cytoplasmic boiling. The characteristics of nuclear fragmentation are particularly visible when THAP-family or THAP domain polypeptide-EGFP condenses in apoptotic bodies.

The ability of the THAP-family- or THAP domain polypeptides to undergo nuclear localization and to induce apoptosis can be tested by transient expression in 293 human kidney cells. If proved susceptible to THAP-family- or THAP domain- induced apoptosis, 293 cells can serve as a convenient initial screen for those THAP family or THAP domain polypeptides, or biologically active fragments or homologues thereof

that will likely also induce apoptosis in other (e.g. endothelial cells or cancer cells). In an exemplary protocol, 293 cells are transfected with plasmid vectors expressing THAP-family- or THAP domain- EGFP fusion protein. Approximately 5* 10^6 293 cells in 100 mm dishes were transfected with 10 g of plasmid DNA using the calcium-phosphate method. The plasmids used are comprise CMV enhancer/promoter and THAP-family- or THAP domain- EGFP coding sequence). Apoptosis is evaluated 24 hrs after transfection by TUNEL and DAPI staining. The THAP-family- or THAP domain- EGFP vector transfected cells are evaluated by fluorescence microscopy with observation of typical nuclear aggregation of the EGFP marker as an indication of apoptosis. If apoptotic, the distribution of EGFP signal in cells expressing THAP-family- or THAP domain-EGFP will be identical to the distribution of DAPI or Hoechst 33342 dyes, which are conventionally used to detect the nuclear DNA changes associated with apoptosis (Cohen et al., supra).

The ability of the THAP family or THAP domain polypeptides, or biologically active fragments or homologues thereof to induce apoptosis can also be tested by expression assays in human cancer cells, for example as available from NCI. Vector type (for example plasmid or retroviral or sindbis viral) can be selected based on efficiency in a given cell type. After the period indicated, cells are evaluated for morphological signs of apoptosis, including aggregation of THAP-family- or THAP domain- EGFP into nuclear apoptotic bodies. Cells are counted under a fluorescence microscope and scored as to the presence or absence of apoptotic signs, or cells are scored by fluorescent TUNEL assay and counted in a flow cytometer. Apoptosis is expressed as a percent of cells displaying typical advanced changes of apoptosis.

Cells from the NCI panel of tumor cells include from example:

-colon cancer, expression using a retroviral expression vector, with evaluation of apoptosis at 96 hrs post-infection (cell lines KM12; HT-29; SW-620; COLO205; HCT-5; HCC 2998; HCT-116);

-CNS tumors, expression using a retroviral expression vector, with evaluation of apoptosis at 96 hrs post-infection (cell lines SF-268, astrocytoma; SF-539, glioblastoma; SNB-19, gliblastoma; SNB-75, astrocytoma; and U251, glioblastoma;

-leukemia cells, expression using a retroviral expression vector, with evaluation of apoptosis at 96 hrs post-infection (cell lines CCRF-CEM, acute lymphocytic

leukemia (ALL); K562, acute myelogenous leukemia (AML); MOLT-4, ALL; SR, immunoblastoma large cell; and RPMI 8226, Myeloblastoma);

-prostate cancer, expression using a retroviral expression vector, with evaluation of apoptosis at 96 hrs post-infection (PC-3);

-kidney cancer, expression using a retroviral expression vector, with evaluation of apoptosis at 96 hrs post-infection (cell lines 768-0; UO-31; TK10; ACHN);

-skin cancer, expression using a retroviral expression vector, with evaluation of apoptosis at 96 hrs post-infection (Melanoma) (cell lines SKMEL-28; M14; SKMEL-5; MALME-3);

-lung cancer, expression using a retroviral expression vector, with evaluation of apoptosis at 96 hrs post-infection (cell lines HOP-92; NCI-H460; HOP-62; NCI-H522; NCI-H23; A549; NCI-H226; EKVX; NCI-H322);

-breast cancer, expression using a retroviral expression vector, with evaluation of apoptosis at 96 hrs post-infection (cell lines MCF-7; T-47D; MCF-7/ADR; MDAMB43; MDAMB23; MDA-N; BT-549);

-ovary cancer, expression using either a retroviral expression vector and protocol or the Sindbis viral expression vector and protocol, with evaluation of apoptosis at 96 hrs post-infection with retrovirus or at 24 hrs post-infection with Sindbis viral vectors (cell lines OVCAR-8; OVCAR-4; IGROV-1; OVCAR-5; OVCAR3; SK-OV-3).

In a further representative example, the susceptibility of malignant melanoma cells to apoptosis induced by a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof can be tested in several known melanoma cell types: human melanoma WM 266-4 (ATCC CRL-1676); human malignant melanoma A-375 (ATCC CRL-1619); human malignant, melanoma A2058 (ATCC CRL-11147); human malignant melanoma SK-MEL-31 (ATCC HTB-73); human malignant melanoma RPMI-7591 ATCC HTB-66 (metastasis to lymph node). Primary melanoma isolates can also be tested. In addition, human chronic myelogenous leukemia K-562 cells (ATCC CCL-243), and 293 human kidney cells (ATCC CRL-1573) (transformed primary embryonal cell) are tested. Normal human primary dermal fibroblasts and Rat-1 fibroblasts serve as controls. All melanoma cell lines are metastatic on the basis of their isolation from metastases or metastatic nodules. A transient expression strategy is used in order to evaluate induction of a THAP-family or

THAP domain polypeptide -mediated apoptosis without artifacts associated with prolonged selection. An expression vector encoding the THAP-family or THAP domain polypeptide -EGFP fusion protein described below can be used in order to facilitate identification of those cells expressing the a THAP-family or THAP domain polypeptide. Cells are transiently transfected by the method optimal for the cell being tested (either $CaPO_4$or Lipofectin). Expression of a THAP-family or THAP domain polypeptide and induction of apoptosis is examined using a fluorescence microscope at 24 hrs and 48 hrs post-transfection. A minimum of approximately 100 cells, which display characteristic EGFP fluorescence, are evaluated by fluorescence microscopy. Apoptosis is scored as nuclear fragmentation, marked apoptotic bodies, and cytoplasmic boiling. The characteristics of nuclear fragmentation are particularly visible when THAP-family or THAP domain polypeptide-EGFP condenses in apoptotic bodies.

In a further example, the susceptibility of endothelial cells to apoptosis induced by a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof can be tested in several known endothelial cell types: HUVEC (human umbilical vein endothelial cells; BioWhittaker-Clonetics, 8830 Biggs Ford Road, Walkersville, MD 21793-0127, Cat N° CC-2519), HMVEC-L (human microvascular endothelial cells from the lung; BioWhittaker-Clonetics, 8830 Biggs Ford Road, Walkersville, MD 21793-0127, Cat N° CC-2527), HMVEC-d (human microvascular endothelial cells from the dermis; BioWhittaker-Clonetics, 8830 Biggs Ford Road, Walkersville, MD 21793-0127, Cat N° CC-2543). These and other endothelial cell types may be useful as models in providing an indication of the ability of THAP-family or THAP domain polypeptides to induce apoptosis in therapeutic strategies for the regulation of angiogenesis. A transient expression strategy is used in order to evaluate induction of a THAP-family or THAP domain polypeptide -mediated apoptosis without artifacts associated with prolonged selection. An expression vector encoding the a THAP-family or THAP domain polypeptide -EGFP fusion protein described below can be used in order to facilitate identification of those cells expressing the a THAP-family or THAP domain polypeptide. Cells are transiently transfected by the method optimal for the cell being tested (either $CaPO_4$ or Lipofectin). Expression of a THAP-family or THAP domain polypeptide and induction of apoptosis is examined using a fluorescence microscope at 24 hrs and 48 hrs post-transfection. A minimum of

approximately 100 cells, which display characteristic EGFP fluorescence, are evaluated by fluorescence microscopy. Apoptosis is scored as nuclear fragmentation, marked apoptotic bodies, and cytoplasmic boiling. The characteristics of nuclear fragmentation are particularly visible when THAP-family or THAP domain polypeptide-EGFP condenses in apoptotic bodies.

In another example, a transient transfection assay procedure is similar to that previously described for detecting apoptosis induced by IL-1-beta-converting enzyme (Miura et al., Cell 75: 653-660 (1993); Kumar et al., Genes Dev. 8: 1613-1626 (1994); Wang et al., Cell 78: 739-750 (1994); and U.S. Patent No. 6,221,615, the disclosures of which are incorporated herein by reference). One day prior to transfection, cells (for example Rat-1 cells) are plated in 24 well dishes at $3.5*10^4$ cells/well. The following day, the cells are transfected with a marker plasmid encoding beta-galactosidase, in combination with an expression plasmid encoding THAP-family or THAP domain polypeptide, by the Lipofectamine procedure (Gibco/BRL). At 24 hours post transfection, cells are fixed and stained with X-Gal to detect beta-galactosidase expression in cells that received plasmid DNA (Miura et al., supra). The number of blue cells is counted by microscopic examination and scored as either live (flat blue cells) or dead (round blue cells). The cell killing activity of the THAP-family or THAP domain polypeptide in this assay is manifested by a large reduction in the number of blue cells obtained relative to co-transfection of the beta-gal plasmid with a control expression vector (i.e., with no THAP-family or THAP domain polypeptide cDNA insert).

In yet another example, beta-galactosidase co-transfection assays can be used for determination of cell death. The assay is performed as described (Hsu, H. et al, (1995). Cell 81,495-504; Hsu, H. et al, (1996a). Cell 84, 299-308; and Hsu, H. et al, (1996b) Inmunity 4, 387-396 and U.S. Patent No. 6,242,569, the disclosures of which are incorporated herein by reference). Transfected cells are stained with X-gal as described in Shu, H. B. et al, ((1995) J. Cell Sci. 108, 2955-2962, the disclosure of which is incorporated herein by reference). The number of blue cells from 8 viewing fields of a 35 mm dish is determined by counting. The average number from one representative experiment is shown.

Assays for apoptosis can also be carried out by making use of any suitable biological marker of apoptosis. Several methods are described as follows.

In one aspect, fluorocytometric studies of cell death status can be carried out. Technology used in fluorocytometric studies employs the identification of cells at three different phases of the cell cycle: G_1, S. and G_2. This is largely performed by DNA quantity staining by propidium iodide labeling. Since the dying cell population contains the same DNA quantity as the living counterparts at any of the three phases of the cell cycle, there is no way to distinguish the two cell populations. One can perform double labeling for a biological marker of apoptosis (e.g. terminin Tp30, U.S. Patent No. 5,783,667) positivity and propidium iodide (PI) staining together. Measurement of the labeling indices for the biological marker of apoptosis and PI staining can be used in combination to obtain the exact fractions of those cells in G_1 that are living and dying. Similar estimations can be made for the S-phase and G_2 phase cell populations.

In this assay, the cells are processed for formaldehyde fixation and extraction with 0.05% Triton. Afterwards, the cell specimens are incubated with monoclonal antibody to a marker of apoptosis overnight at room temperature or at 37C for one hour. This is followed by further incubation with fluoresceinated goat antimouse antibody, and subsequent incubation by propidium iodide staining. The completely processed cell specimens are then evaluated by fluorocytometric measurement on both fluorescence (marker of apoptosis) and rhodamine (PI) labeling intensity on a per cell basis, with the same cell population simultaneously.

In another aspect, it is possible to assess the inhibitory effect on cell growth by therapeutic induction of apoptosis. One routine method to determine whether a particular chemotherapeutic drug can inhibit cancerous cell growth is to examine cell population size either in culture, by measuring the reduction in cell colony size or number, or measuring soft agar colony growth or in vivo tumor formation in nude mice, which procedures require time for development of the colonies or tumor to be large enough to be detectable. Experiments involved in these approaches in general require large-scale planning and multiple repeats of lengthy experimental span (at least three weeks). Often these assays do not take into account the fact that a drug may not be inhibiting cell growth, but rather killing the cells, a more favorable consequence needed for chemotherapeutic treatment of cancer. Thus, assays for the assessment of apoptotis activity can involve using a biological or biochemical marker specific for quiescent, non-cycling or non-proliferating cells. For example, a monoclonal antibody can be used

to assess the non-proliferating population of cells in a given tissue which indirectly gives a measure of the proliferating component of a tumor or cell mass. This detection can be combined with a biological or biochemical marker (e.g. antibodies) to detect the dying cell population pool, providing a powerful and rapid assessment of the effectiveness of any given drugs in the containment of cancerous cell growth. Applications can be easily performed at the immunofluorescence microscopic level with cultured cells or tissue sections.

In other aspects, a biological or biochemical marker can be used to assess pharmacological intervention on inhibition of cell death frequency in degenerative diseases. For degenerative diseases such as Alzheimer's or Parkinson's disease, these losses may be due to the premature activation of the cell death program in neurons. In osteoporosis, the cell loss may be due to an improper balance between osteoblast and osteoclast cells, due to the too active programmed cell death process killing more cells than the bone tissue can afford. Other related phenomena may also occur in the wound healing process, tissue transplantation and cell growth in the glomerus during kidney infection, where the balance between living and dying cell populations is an essential issue to the health status of the tissue, and are further described in the section titled "Methods of treatment". A rapid assessment of dying cell populations can be made through the immunohistochemical and biochemical measurements of a biological or biochemical marker of apoptosis in degenerative tissues. In one example, a biological or biochemical marker can be used to assess cell death status in oligodendrocytes associated with Multiple Sclerosis. Positive staining of monoclonal antibody to a marker of apoptosis (such as Tp30, U.S. Patent No. 5,783,667, the disclosure of which is incorporated herein by reference) occurs in dying cultured human oligodendrocytes. The programmed cell death event is activated in these oligodendrocytes by total deprivation of serum, or by treatment with tumor necrosis factor (TNF).

In general, a biological or biochemical marker can also be used to assess cell death status in pharmacological studies in animal models. Attempting to control either a reduced cell death rate, in the case of cancer, or an increased cell death rate, in the case of neurodegeneration, has been recently seen as a new mode of disease intervention. Numerous approaches via either intervention with known drugs or gene therapy are in progress, starting from the base of correcting the altered programmed cell

death process, with the concept on maintaining a balanced cell mass in any given tissue. For these therapeutic interventions, the bridge between studies in cultured cells and clinical trials is animal studies, i.e. success in intervention with animal models, in either routine laboratory animals or transgenic mice bearing either knock-out or overexpression phenotypes. Thus, a biological or biochemical marker of apoptosis, such as an antibody for an apoptosis-specific protein, is a useful tool for examining apoptotic death status in terms of change in dying cell numbers between normal and experimentally manipulated animals. In this context the invention, as a diagnostic tool for assessing cell death status, could help to determine the efficacy and potency of a drug or a gene therapeutic approach.

As discussed, provided are methods for assessing the activity of THAP-family members and therapeutic treatment acting on THAP-family members or related biological pathways. However, in other aspects, the same methods may be used for assessment of apoptosis in general, when a THAP-family member is used as a biological marker of apoptosis. Thus, the invention also provides diagnostic and assay methods using a THAP-family member as a marker of cell death or apoptotic activity. Further diagnostic assays are also provided herein in the section titled 'Diagnostic and prognostic uses'.

CHEMOKINE BINDING BY THAP-FAMILY PROTEINS

Some embodiments of the present invention relate to THAP-family polypeptides, chemokine-binding domains of THAP-family polypeptides, THAP oligomers, and chemokine-binding domain-THAP-immunoglobulin fusion proteins such as those described above which bind to chemokines other than SLC. For example, THAP-family polypeptides, chemokine-binding domains of THAP-family polypeptides, THAP oligomers, and chemokine-binding domain-THAP-immunoglobulin fusion proteins can be used to bind to or otherwise interact with chemokines from many families such as C chemokines, CC chemokines, C-X-C chemokines, C-X3-C chemokines, XC chemokines or CCK chemokines. In particular, THAP-family polypeptides, chemokine-binding domains of THAP-family polypeptides, THAP oligomers, and chemokine-binding domain-THAP-immunoglobulin fusion proteins may interact with chemokines such as XCL1, XCL2, CCL1, CCL2, CCL3, CCL3L1,

SCYA3L2, CCL4, CCL4L, CCL5, CCL6, CCL7, CCL8, SCYA9, SCYA10, CCL11, SCYA12, CCL13, CCL14, CCL15, CCL16, CCL17, CCL18, CCL19, CCL20, CCL21, CCL22, CCL23, CCL24, CCL25, CCL26, CCL27, CCL28, clone 391, CARP CC-1, CCL1, CK-1, regakine-1, K203, CXCL1, CXCL1P, CXCL2, CXCL3, PF4, PF4V1, CXCL5, CXCL6, PPBP, SPBPBP, IL8, CXCL9, CXCL10, CXCL11, CXCL12, CXCL14, CXCL15, CXCL16, NAP-4, LFCA-1, Scyba, JSC, VHSV-induced protein, CX3CL1 and fCL1.

In some embodiments of the present invention, THAP-family polypeptides, chemokine-binding domains of THAP-family polypeptides, THAP oligomers, and chemokine-binding domain-THAP-immunoglobulin fusion proteins can bind to a chemokine extracellularly. For example, the interaction of the THAP1 polypeptide, a biologically active fragment thereof (such as the chemokine-binding domain of THAP1 (amino acids 143-213 of SEQ ID NO: 3)), an oligomer thereof, or an immunoglobulin fusion thereof can bind to a chemokine extracellularly. In other examples, chemokine-binding domains of other THAP-family members such as THAP2, THAP3, THAP4, THAP5, THAP6, THAP7, THAP8, THAP9, THAP10, THAP11 or THAP0, biologically active fragments thereof, oligomers thereof, or immunoglobulin fusions thereof can be used to bind to chemokines extracellularly. Binding of the THAP-family polypeptides, chemokine-binding domains of THAP-family polypeptides, THAP oligomers, and chemokine-binding domain-THAP-immunoglobulin fusion proteins may either decrease or increase the affinity of the chemokine for its extracellular receptor. In cases where binding of the chemokine to its extracellular receptor is inhibited, the normal biological effect of the chemokine can be inhibited. Such inhibition can prevent the occurrence of chemokine-mediated cellular responses, such as the modulation of cell proliferation, the modulation of angiogenesis, the modulation of an inflammation response, the modulation of apoptosis, the modulation of cell differentiation. Alternatively, in cases where binding of the chemokine to its extracellular receptor is activated, the normal biological effect of the chemokine can be enhanced. Such enhancement can increase the occurrence of chemokine-mediated cellular responses, such as the modulation of cell proliferation, the modulation of angiogenesis, the modulation of an inflammation response, the modulation of apoptosis, the modulation of cell differentiation.

As used herein, "ELC/CCL19", "CCL19" and "ELC" are synonymous.

As used herein, "Rantes/CCL5", "CCL5" and "Rantes" are synonymous.

As used herein, "MIG/CXCL9", "CXCL9" and "MIG" are synonymous.

As used herein, "IP10/CXCL10", "CXCL10" and "IP10" are synonymous.

In some embodiments of the present invention a chemokine-binding domain that consists essentially of the chemokine binding portion of a THAP-family polypeptide is contemplated. In some embodiments, the THAP-family polypeptide is THAP-1 (SEQ ID NO: 3) or a homolog thereof. Chemokines that are capable of binding to any particular THAP-family member can be determined as described in Examples 16, 32 and 33, which set out both *in vitro* and *in vivo* assays for determining the binding affinity of several different chemokines to THAP-1. The portion of the THAP-family protein that binds to the chemokine can readily be determined through the analysis of deletion and point mutants of any of the THAP-family members capable of chemokine-binding. Such analyses of deletion and point mutants were used to determine the specific region of THAP-1 that permits SLC-binding (see Example 15). Additionally, deletion and point mutation studies were used to determine portions of THAP-family proteins as well as specific amino acid residues that interact with PAR-4 (Examples 4-7 and 13). It will be appreciated that the methods described in these Examples can be used to precisely identify the chemokine-binding portion of any THAP-family member using any chemokine.

By "chemokine-binding domain" or "portion that binds to a chemokine" is meant a fragment which comprises 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 160, 170, 180, 190, 200, 210 or greater than 210 consecutive amino acids of a THAP-family polypeptide but less than the total number of amino acids present in the THAP-family polypeptide. In some embodiments, the THAP-family polypeptide is THAP-1 (SEQ ID NO: 3).

The complete amino acid sequence of each human THAP-family polypeptide is described in the Sequence Listing. In particular, THAP-1 is (SEQ ID NO: 3), THAP-2 is (SEQ ID NO: 4), THAP-3 is (SEQ ID NO: 5), THAP-4 is (SEQ ID NO: 6), THAP-5 is (SEQ ID NO: 7), THAP-6 is (SEQ ID NO: 8), THAP-7 is (SEQ ID NO: 9), THAP-8 is (SEQ ID NO: 10), THAP-9 is (SEQ ID NO:11), THAP-10 is (SEQ ID NO: 12), THAP-11 is (SEQ ID NO: 13), THAP-0 is (SEQ ID NO: 14). The complete amino acid sequence of additional THAP-family polypeptides from other species are also listed in the Sequence Listing as SEQ ID NOs: 16-98. As such, the chemokine-binding portion of any of these THAP-family polypeptide sequences that are listed in the Sequence Listing is explicitly described. In particular, in some embodiments, the chemokine-binding domain is a fragment of a THAP-family chemokine-binding agent described by the formula:

for each THAP-family polypeptide, N = the number of amino acids in the full-length polypeptide; B = a number between 1 and N – 1; and E = a number between 1 and N.

For any THAP-family polypeptide, a chemokine-binding domain is specified by any consecutive sequence of amino acids beginning at an amino acid position B and ending at amino acid position E, wherein E > B.

Methods Of Complex Formation Between A Chemokine And A THAP-Type Chemokine-Binding Agent

Some aspects of the present invention related to methods for forming a complex between a chemokine and a THAP-type chemokine-binding agent. These methods include the step of contacting one or more chemokines with one or more THAP-type chemokine-binding agents described herein such that a complex comprising one or more chemokines and one or more THAP-type chemokine-binding agents is formed. In some embodiments, a plurality of different chemokines are contacted with one or a plurality of different THAP-type chemokine-binding agents so as to form one or more complexes. Alternatively, a plurality of different THAP-type chemokine-binding agents are contacted with one or more chemokines so as to form one or more complexes.

A number of different chemokines can be used in the above-described complex

formation methods. Such chemokines include, but are not limited to, XCL1, XCL2, CCL1, CCL2, CCL3, CCL3L1, SCYA3L2, CCL4, CCL4L, CCL5, CCL6, CCL7, CCL8, SCYA9, SCYA10, CCL11, SCYA12, CCL13, CCL14, CCL15, CCL16, CCL17, CCL18, CCL19, CCL20, CCL21, CCL22, CCL23, CCL24, CCL25, CCL26, CCL27, CCL28, clone 391, CARP CC-1, CCL1, CK-1, regakine-1, K203, CXCL1, CXCL1P, CXCL2, CXCL3, PF4, PF4V1, CXCL5, CXCL6, PPBP, SPBPBP, IL8, CXCL9, CXCL10, CXCL11, CXCL12, CXCL14, CXCL15, CXCL16, NAP-4, LFCA-1, Scyba, JSC, VHSV-induced protein, CX3CL1 and fCL1.

Method of forming a complex between a THAP-type chemokine-binding agent and a chemokine can be used both *in vitro* and *in vivo*. For example, *in vitro* uses can include the detection of a chemokine in a solution or a biological sample that has been removed or withdrawn from a subject. Such samples may include, but are not limited to, tissue samples, blood samples, and other fluid or solid samples of biological material. *In vivo* uses can include, but are not limited to, the detection or localization of chemokines in a subject, reducing or inhibiting the activity of one or more chemokines throughout or in certain areas of a subject's body, and reducing the symptoms associated with a chemokine influenced or mediated condition.

METHODS OF TREATMENT

A large body of evidence gathered from experiments carried out with apoptosis modulating strategies suggests that treatments acting on apoptosis-inducing or cell proliferation-reducing proteins may offer new treatment methods for a wide range of disorders. Methods of treatment according to the invention may act in a variety of manners, given the novel function provided for a number of proteins, and the linking of several biological pathways.

Provided herein are treatment methods based on the functionalization of the THAP-family members. THAP family or THAP domain polypeptides, and biologically active fragments and homologues thereof, as described further herein may be useful in modulation of apoptosis or cell proliferation.

The methods of treatment involve acting on a molecule of the invention (that is, a THAP family member polypeptide, THAP-family target, or PAR4 or PAR4 target). Included are methods which involve modulating THAP-family polypeptide activity,

THAP-family target activity, or PAR4 or PAR4 target activity. This modulation (increasing or decreasing) of activity can be carried out in a number of suitable ways, several of which have been described in the present application.

For example, methods of treatment may involve modulating a "THAP-family activity", "biological activity of a THAP-family member" or "functional activity of a THAP-family member". Modulating THAP-family activity may involve modulating an association with a THAP-family-target molecule (for example, association of THAP1, THAP2 or THAP3 with Par4 or association of THAP1, THAP2 or THAP3 with a PML-NB protein) or preferably any other activity selected from the group consisting of: (1) mediating apoptosis or cell proliferation when expressed or introduced into a cell, most preferably inducing or enhancing apoptosis, and/or most preferably reducing cell proliferation; (2) mediating apoptosis or cell proliferation of an endothelial cell; (3) mediating apoptosis or cell proliferation of a hyperproliferative cell; (4) mediating apoptosis or cell proliferation of a CNS cell, preferably a neuronal or glial cell; or (5) an activity determined in an animal selected from the group consisting of mediating, preferably inhibiting angiogenesis, mediating, preferably inhibiting inflammation, inhibition of metastatic potential of cancerous tissue, reduction of tumor burden, increase in sensitivity to chemotherapy or radiotherapy, killing a cancer cell, inhibition of the growth of a cancer cell, or induction of tumor regression. Detecting THAP-family activity may also comprise detecting any suitable therapeutic endpoint associated with a disease condition discussed herein.

In another example, methods of treatment may involve modulating a "PAR4 activity", "biological activity of PAR4" or "functional activity of PAR4 ". Modulating PAR4 activity may involve modulating an association with a PAR4-target molecule (for example THAP1, THAP2, THAP3 or PML-NB protein) or most preferably PAR4 apoptosis inducing or enhancing (e.g. signal transducing) activity, or inhibition of cell proliferation or cell cycle.

Methods of treatment may involve modulating the recruitment, binding or association of proteins to PML-NBs, or otherwise modulating PML-NBs activity. The present invention also provides methods for modulating PAR4 activity, comprising modulating PAR4 interactions with THAP-family proteins, and PAR4 and PML-NBs, as well as modulating THAP-family activity, comprising modulating for example

THAP1 interactions with PML-NBs. The invention encompasses inhibiting or increasing the recruitment of THAP1, or PAR4 to PML-NBs. Preventing the binding of either or both of THAP1 or PAR4 to PML-NBs may increase the bioavailability or THAP1 and/or PAR4, thus providing a method of increasing THAP1 and/or PAR4 activity. The invention also encompasses inhibiting or increasing the binding of a THAP-family protein (such as THAP1) or PAR4 to PML-NBs or to another protein associated with PML-NBs, such as a protein selected from the group consisting of daxx, sp100, sp140, p53, pRB, CBP, BLM, SUMO-1. For example, the invention encompasses modulating PAR4 activity by preventing the binding of THAP1 to PAR4, or by preventing the recruitment or binding of PAR4 to PML-NBs.

Therapeutic methods and compositions of the invention may involve (1) modulating apoptosis or cell proliferation, most preferably inducing or enhancing apoptosis, and/or most preferably reducing cell proliferation; (2) modulating apoptosis or cell proliferation of an endothelial cell (3) modulating apoptosis or cell proliferation of a hyperproliferative cell; (4) modulating apoptosis or cell proliferation of a CNS cell, preferably a neuronal or glial cell; (5) inhibition of metastatic potential of cancerous tissue, reduction of tumor burden, increase in sensitivity to chemotherapy or radiotherapy, killing a cancer cell, inhibition of the growth of a cancer cell, or induction tumor regression; or (6) interaction with a THAP family target molecule or THAP domain target molecule, preferably interaction with a protein or a nucleic acid. Methods may also involve improving a symptom of or ameliorating a condition as further described herein.

Antiapoptotic therapy

Molecules of the invention (e.g. those obtained using the screening methods described herein, dominant negative mutants, antibodies etc.) which inhibit apoptosis are also expected to be useful in the treatment and/or prevention of disease. Diseases in which it is desirable to prevent apoptosis include neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, retinitis pigmentosa and cerebellar degeneration; myelodysplasis such as aplastic anemia; ischemic diseases such as myocardial infarction and stroke; hepatic diseases such as alcoholic hepatitis, hepatitis B and hepatitis C; joint-diseases such as osteoarthritis;

atherosclerosis; and etc. The apoptosis inhibitor of the present invention is especially preferably used as an agent for prophylaxis or treatment of a neurodegenerative disease (see also Adams, J. M., Science, 281:1322 (1998).

Included as inhibitors of apoptosis as described herein are generally any molecule which inhibits activity of a THAP family or THAP domain polypeptide, or a biologically active fragment or homologue thereof, a THAP-family target protein or PAR4 (particularly PAR4/PML-NB protein interactions). THAP-family and THAP domain polypeptides inhibitors may include for example antibodies, peptides, dominant negative THAP-family or THAP domain analogs, small molecules, ribozyme or antisense nucleic acids. These inhibitors may be particularly advantageous in the treatment of neurodegenerative disorders. Particularly preferred are inhibitors which affect binding of THAP-family protein to a THAP-family target protein, and inhibitors which affect the DNA binding activity of a THAP-family protein.

In further preferred aspects the invention provides inhibitors of THAP-family activity, including but not limited to molecules which interfere or inhibit interactions of THAP-family proteins with PAR4, for the treatment of endothelial cell related disorders and neurodegenerative disorders. Support is found in the literature, as PAR4 appears to play a key role in neuronal apoptosis in various neurodegenerative disorders (Guo et al., 1998; Mattson et al., 2000; Mattson et al., 1999; Mattson et al., 2001). THAP1, which is expressed in brain and associates with PAR4 may therefore also play a key role in neuronal apoptosis. Drugs that inhibit THAP-family and/or inhibit THAP-family/PAR4 complex formation may lead to the development of novel preventative and therapeutic strategies for neurodegenerative disorders.

Apoptosis regulation in endothelial cells

The invention also provides methods of regulating angiogenesis in a subject which are expected to be useful in the treatment of cancer, cardiovascular diseases and inflammatory diseases. An inducer of apoptosis of immortalized cells is expected to be useful in suppressing tumorigenesis and/or metastasis in malignant tumors. Examples of malignant tumors include leukemia (for example, myelocytic leukemia, lymphocytic leukemia such as Burkitt lymphoma), digestive tract carcinoma, lung carcinoma, pancreas carcinoma, ovary carcinoma, uterus carcinoma, brain tumor, malignant

melanoma, other carcinomas, and sarcomas. The present inventors have isolated both THAP1 and PAR4 cDNAs from human endothelial cells, and both PAR4 and PML are known to be expressed predominantly in blood vessel endothelial cells (Boghaert et al., (1997) Cell Growth Differ 8(8):881-90; Terris B. et al, (1995) Cancer Res. 55(7):1590-7, 1995, the disclosures of which are incorporated herein by reference), suggesting that the PML-NBs-and the newly associated THAP1/PAR4 proapoptotic complex may be a major regulator of endothelial cell apoptosis *in vivo* and thus constitute an attractive therapeutic target for angiogenesis-dependent diseases. For example, THAP1 and PAR4 pathways may allow selective treatments that regulate (e.g. stimulate or inhibit) angiogenesis.

In a first aspect, the invention provides methods of inhibiting endothelial cell apoptosis, by administering a THAP1 or PAR4 inhibitor, or optionally a THAP1/PAR4 interaction inhibitor or optionally an inhibitor of THAP1 DNA binding activity. As further described herein, the THAP domain is involved in THAP1 pro-apoptotic activity. Deletion of the THAP domain abrogates the proapoptotic activity of THAP1 in mouse 3T3 fibroblasts, as shown in Example 11. Also, as further described herein, deletion of residues 168-172 or replacement of residues 171-172 abrogates THAP1 binding to PAR4 both in vitro and in vivo and results in lack of recruitment of PAR4 by THAP1 to PML-NBs. For PAR4, the leucine zipper domain is required (and is sufficient) for binding to THAP1.

Inhibiting endothelial cell apoptosis may improve angiogenesis and vasculogenesis in patients with ischemia and may also interfere with focal dysregulated vascular remodeling, the key mechanism for atherosclerotic disease progression.

In another aspect, the invention provides methods of inducing endothelial cell apoptosis, by administering for example a biologically active THAP family polypeptide such as THAP1, a THAP domain polypeptide or a PAR4 polypeptide, or a biologically active fragment or homologue thereof, or a THAP1 or PAR4 stimulator. Stimulation of endothelial cell apoptosis may prevent or inhibit angiogenesis and thus limit unwanted neovascularization of tumors or inflamed tissues (see Dimmeler and Zeiher, Circulation Research, 2000, 87 :434-439, the disclosure of which is incorporated herein by reference).

Angiogenesis

Angiogenesis is defined in adult organism as the formation of new blood vessels by a process of sprouting from pre-existing vessels. This neovascularization involves activation, migration, and proliferation of endothelial cells and is driven by several stimuli, among those shear stress. Under normal physiological conditions, humans or animals undergo angiogenesis only in very specific restricted situations. For example, angiogenesis is normally observed in wound healing, fetal and embryonal development and formation of the corpus luteum, endometrium and placenta. Molecules of the invention may have endothelial inhibiting or inducing activity, having the capability to inhibit or induce angiogenesis in general.

Both controlled and uncontrolled angiogenesis are thought to proceed in a similar manner. Endothelial cells and pericytes, surrounded by a basement membrane, form capillary blood vessels. Angiogenesis begins with the erosion of the basement membrane by enzymes released by endothelial cells and leukocytes. The endothelial cells, which line the lumen of blood vessels, then protrude through the basement membrane. Angiogenic stimulants induce the endothelial cells to migrate through the eroded basement membrane. The migrating cells form a "sprout" off the parent blood vessel, where the endothelial cells undergo mitosis and proliferate. The endothelial sprouts merge with each other to form capillary loops, creating the new blood vessel.

Persistent, unregulated angiogenesis occurs in a multiplicity of disease states, tumor metastasis and abnormal growth by endothelial cells and supports the pathological damage seen in these conditions. The diverse pathological disease states in which unregulated angiogenesis is present have been grouped together as angiogenic dependent or angiogenic associated diseases. It is thus an object of the present invention to provide methods and compositions for treating diseases and processes that are mediated by angiogenesis including, but not limited to, hemangioma, solid tumors, leukemia, metastasis, telangiectasia psoriasis scleroderma, pyogenic granuloma, Myocardial angiogenesis, plaque neovascularization, cororany collaterals, ischemic limb angiogenesis, corneal diseases, rubeosis, neovascular glaucoma, diabetic retinopathy, retrolental fibroplasia, arthritis, diabetic neovascularization, macular degeneration, wound healing, peptic ulcer, fractures, keloids, vasculogenesis, hematopoiesis, ovulation, menstruation, and placentation.

(i) Anti-angiogenic therapy

In one aspect the invention provides anti-angiogenic therapies as potential treatments for a wide variety of diseases, including cancer, arteriosclerosis, obesity, arthritis, duodenal ulcers, psoriasis, proliferative skin disorders, cardiovascular disorders and abnormal ocular neovascularization caused, for example, by diabetes (Folkman, Nature Medicine 1:27 (1995) and Folkman, Seminars in Medicine of the Beth Israel Hospital, Boston, New England Journal of Medicine, 333:1757 (1995)). Anti-angiogenic therapies are thought to act by inhibiting the formation of new blood vessels.

The present invention thus provides methods and compositions for treating diseases and processes mediated by undesired and uncontrolled angiogenesis by administering to a human or animal a composition comprising a substantially purified THAP family or THAP domain polypeptide, or a biologically active fragment, homologue or derivative thereof in a dosage sufficient to inhibit angiogenesis, administering a vector capable of expressing a nucleic acid encoding a THAP-family or THAP domain protein, or administering any other inducer of expression or activity of a THAP-family or THAP domain protein. The present invention is particularly useful for treating or for repressing the growth of tumors. Administration of THAP-family or THAP domain nucleic acid, protein or other inducer to a human or animal with prevascularized metastasized tumors will prevent the growth or expansion of those tumors. THAP-family activity may be used in combination with other compositions and procedures for the treatment of diseases. For example, a tumor may be treated conventionally with surgery, radiation or chemotherapy combined with THAP-family or THAP domain protein and then THAP-family or THAP domain protein may be subsequently administered to the patient to extend the dormancy of micrometastases and to stabilize any residual primary tumor.

In a preferred example, a THAP-family polypeptide activity, preferably a THAP1 activity is used for the treatment of arthritis, for example rheumatiod arthritis. Rheumatoid arthritis is characterized by symmetric, polyarticular inflammation of synovial-lined joints, and may involve extraarticular tissues, such as the pericardium, lung, and blood vessels.

(ii) Angiogenic therapy

In another aspect, the inhibitors of THAP-family protein activity, particularly THAP1 activity, could be used as an anti-apoptotic and thus as an angiogenic therapy. Angiogenic therapies are potential treatments for promoting wound healing and for stimulating the growth of new blood vessels to by-pass occluded ones. Thus, pro-angiogenic therapies could potentially augment or replace by-pass surgeries and balloon angioplasty (PTCA). For example, with respect to neovascularization to bypass occluded blood vessels, a "therapeutically effective amount" is a quantity which results in the formation of new blood vessels which can transport at least some of the blood which normally would pass through the blocked vessel.

The THAP-family protein of the present invention can for example be used to generate antibodies that can be used as inhibitors of apoptosis. The antibodies can be either polyclonal antibodies or monoclonal antibodies. In addition, these antibodies that specifically bind to the THAP-family protein can be used in diagnostic methods and kits that are well known to those of ordinary skill in the art to detect or quantify the THAP-family protein in a body fluid. Results from these tests can be used to diagnose or predict the occurrence or recurrence of a cancer and other angiogenic mediated diseases.

It will be appreciated that other inhibitors of THAP-family and THAP domain proteins can also be used in angiogenic therapies, including for example small molecules, antisense nucleic acids, dominant negative THAP-family and THAP domain proteins or peptides identified using the above methods.

In view of applications in both angiogenic and antiangiogenic therapies, molecules of the invention may have endothelial inhibiting or inducing activity, having the capability to inhibit or induce angiogenesis in general. It will be appreciated that methods of assessing such capability are known in the art, including for example assessing antiangiogenic properties as the ability inhibit the growth of bovine capillary endothelial cells in culture in the presence of fibroblast growth factor.

It is to be understood that the present invention is contemplated to include any derivatives of the THAP family or THAP domain polypeptides, and biologically active fragments and homologues thereof that have endothelial inhibitory or apoptotic activity. The present invention includes full-length THAP-family and THAP domain proteins,

derivatives of the THAP-family and THAP domain proteins and biologically-active fragments of the THAP-family and THAP domain proteins. These include proteins with THAP-family protein activity that have amino acid substitutions or have sugars or other molecules attached to amino acid functional groups. The methods also contemplate the use of genes that code for a THAP-family protein and to proteins that are expressed by those genes.

As discussed, several methods are described herein for delivering a modulator to a subject in need of treatment, including for example small molecule modulators, nucleic acids including via gene therapy vectors, and polypeptides including peptide mimetics, active polypeptides, dominant negative polypeptides and antibodies. It will be thus be appreciated that modulators of the invention identified according to the methods in the section titled "Drug Screening Assays" can be further tested in cell or animal models for their ability to ameliorate or prevent a condition involving a THAP-family polypeptide, particularly THAP1, THAP1, THAP2 or THAP3/PAR4 interactions, THAP-family DNA binding or PAR4 / PML-NBs interactions. Likewise, nucleic acids, polypeptides and vectors (e.g. viral) can also be assessed in a similar manner.

An "individual" treated by the methods of this invention is a vertebrate, particularly a mammal (including model animals of human disease, farm animals, sport animals, and pets), and typically a human. "Individual" is also synonymous with "subject."

"Treatment" refers to clinical intervention in an attempt to alter the natural course of the individual being treated, and may be performed either for prophylaxis or during the course of clinical pathology. Desirable effects include preventing occurrence or recurrence of disease, alleviation of symptoms, diminishment of any direct or indirect pathological consequences of the disease, such as hyperresponsiveness, inflammation, or necrosis, lowering the rate of disease progression, amelioration or palliation of the disease state, and remission or improved prognosis. The "pathology" associated with a disease condition is anything that compromises the well-being, normal physiology, or quality of life of the affected individual.

Treatment is performed by administering an effective amount of a THAP-family polypeptide inhibitor or activator. An "effective amount" is an amount sufficient to effect a beneficial or desired clinical result, and can be administered in one or more doses. The criteria for assessing response to therapeutic modalities employing the lipid compositions of this invention are dictated by the specific condition, measured according to standard medical procedures appropriate for the condition.

REDUCING CHEMOKINE MEDIATED EFFECTS

Some aspects of the present invention relate to the use of THAP-family polypeptides, including THAP-1, chemokine-binding domains of THAP-family polypeptides, THAP-family polypeptide or THAP-family chemokine-binding domain fusions to immunoglobulin Fc, oligomers of THAP-family polypeptides or THAP-family chemokine-binding domains, or homologs of any of the above-listed compositions (together and herein after referred to as THAP-type chemokine-binding agents) for reducing the inflammation or the symptoms associated with diseases or conditions that are influenced or mediated by chemokine binding or activity. In such embodiments, the THAP-type chemokine binding agents are administered to a subject in effective amounts so as to reduce the symptoms associated with the condition. In some embodiments, the chemokine that is effected by the THAP-type chemokine binding agent is SLC, CCL19, CCL5, CXCL9, CXCL10 or a combination of these chemokines. In other embodiments, the chemokine that is effected by the THAP-type chemokine binding agent is XCL1, XCL2, CCL1, CCL2, CCL3, CCL3L1, SCYA3L2, CCL4, CCL4L, CCL5, CCL6, CCL7, CCL8, SCYA9, SCYA10, CCL11, SCYA12, CCL13, CCL14, CCL15, CCL16, CCL17, CCL18, CCL19, CCL20, CCL21, CCL22, CCL23, CCL24, CCL25, CCL26, CCL27, CCL28, clone 391, CARP CC-1, CCL1, CK-1, regakine-1, K203, CXCL1, CXCL1P, CXCL2, CXCL3, PF4, PF4V1, CXCL5, CXCL6, PPBP, SPBPBP, IL8, CXCL9, CXCL10, CXCL11, CXCL12, CXCL14, CXCL15, CXCL16, NAP-4, LFCA-1, Scyba, JSC, VHSV-induced protein, CX3CL1, fCL1 or a combination of these chemokines. In some embodiments, the THAP-type chemokine-binding agent is administered directly whereas in other embodiments it is administered as a pharmaceutical composition. In either case, the routes of administration that are known in the art and described herein may be used to deliver the THAP-type chemokine-binding agent to the subject.

Some embodiments of the present invention relate to a device for delivering the THAP-type chemokine-binding agent or pharmaceutical composition thereof to the subject. In such embodiment, the device comprises a container which contains the THAP-type chemokine-binding agent or pharmaceutical composition thereof. For example, in some embodiments, the device may be a conventional device including, but not limited to, syringes, devices for intranasal administration of compositions and vaccine guns. In one embodiment, the device comprises a member which receives the THAP-type chemokine-binding agent or pharmaceutical composition thereof in communication with a mechanism for delivering the composition to the subject. In some embodiments, the device is an inhaler or a patch for transdermal administration.

Pharmaceutical Compositions

Compounds capable of inhibiting THAP-family activity, preferably small molecules but also including peptides, THAP-family nucleic acid molecules, THAP-family proteins, and anti-THAP-family antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents;

antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor ELα (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Where the active compound is a protein, peptide or anti-THAP-family antibody, sterile injectable solutions can be prepared by incorporating the active compound (e.g.,) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle

which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer. Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art. Most preferably, active compound is delivered to a subject by intravenous injection.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Pat. No. 4,522,811, the disclosure of which is incorporated herein by reference in its entirety.

It is especially advantageous to formulate oral or preferably parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

It will be appreciated that THAP-type chemokine-binding agents can be formulated as pharmaceutical compositions and administered as described above. Additionally, the effective dose, route of administration, duration of administration, duration between doses and therapeutic effect can be determined by the methods described above as well as using methods that are well known in the art.

Diagnostic and Prognostic Uses

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics; and in drug screening and methods of treatment (e.g., therapeutic and prophylactic) as further described herein.

The invention provides diagnostic and prognositc assays for detecting THAP-family members, as further described. Also provided are diagnostic and prognostic assays for detecting interactions between THAP-family members and THAP-family target molecules. In a preferred example, a THAP-family member is THAP1, THAP2 or THAP3 and the THAP-family target is PAR4 or a PML-NB protein.

The invention also provides diagnostic and prognositc assays for detecting THAP1 and/or PAR4 localization to or association with PML-NBs, or association with or binding to a PML-NB-associated protein, such as daxx, sp100, sp140, p53, pRB, CBP, BLM or SUMO-1. In a preferred method, the invention provides detecting PAR4 localization to or association with PML-NBs. In a further aspect, the invention provides detecting THAP-family nucleic acid binding activity.

The isolated nucleic acid molecules of the invention can be used, for example, to detect THAP-family polypeptide mRNA (e.g., in a biological sample) or a genetic alteration in a THAP-family gene, and to modulate a THAP-family polypeptide activity, as described further below. The THAP-family proteins can be used to treat disorders characterized by insufficient or excessive production of a THAP-family protein or THAP-family target molecules. In addition, the THAP-family proteins can be used to screen for naturally occurring THAP-family target molecules, to screen for drugs or

compounds which modulate, preferably inhibit THAP-family activity, as well as to treat disorders characterized by insufficient or excessive production of THAP-family protein or production of THAP-family protein forms which have decreased or aberrant activity compared to THAP-family wild type protein. Moreover, the anti-THAP-family antibodies of the invention can be used to detect and isolate THAP-family proteins, regulate the bioavailability of THAP-family proteins, and modulate THAP-family activity.

Accordingly one embodiment of the present invention involves a method of use (e.g., a diagnostic assay, prognostic assay, or a prophylactic/therapeutic method of treatment) wherein a molecule of the present invention (e.g., a THAP-family protein, THAP-family nucleic acid, or most preferably a THAP-family inhibitor or activator) is used, for example, to diagnose, prognose and/or treat a disease and/or condition in which any of the aforementioned THAP-family activities is indicated. In another embodiment, the present invention involves a method of use (e.g., a diagnostic assay, prognostic assay, or a prophylactic/therapeutic method of treatment) wherein a molecule of the present invention (e.g., a THAP-family protein, THAP-family nucleic acid, or a THAP-family inhibitor or activator) is used, for example, for the diagnosis, prognosis, and/or treatment of subjects, preferably a human subject, in which any of the aforementioned activities is pathologically perturbed. In a preferred embodiment, the methods of use (e.g., diagnostic assays, prognostic assays, or prophylactic/therapeutic methods of treatment) involve administering to a subject, preferably a human subject, a molecule of the present invention (e.g., a THAP-family protein, THAP-family nucleic acid, or a THAP-family inhibitor or activator) for the diagnosis, prognosis, and/or therapeutic treatment. In another embodiment, the methods of use (e.g., diagnostic assays, prognostic assays, or prophylactic/therapeutic methods of treatment) involve administering to a human subject a molecule of the present invention (e.g., a THAP-family protein, THAP-family nucleic acid, or a THAP-family inhibitor or activator).

For example, the invention encompasses a method of determining whether a THAP-family member is expressed within a biological sample comprising: a) contacting said biological sample with: ii) a polynucleotide that hybridizes under stringent conditions to a THAP-family nucleic acid; or iii) a detectable polypeptide (e.g. antibody) that selectively binds to a THAP-family polypeptide; and b) detecting the

presence or absence of hybridization between said polynucleotide and an RNA species within said sample, or the presence or absence of binding of said detectable polypeptide to a polypeptide within said sample. A detection of said hybridization or of said binding indicates that said THAP-family member is expressed within said sample. Preferably, the polynucleotide is a primer, and wherein said hybridization is detected by detecting the presence of an amplification product comprising said primer sequence, or the detectable polypeptide is an antibody.

Also envisioned is a method of determining whether a mammal, preferably human, has an elevated or reduced level of expression of a THAP-family member, comprising: a) providing a biological sample from said mammal; and b) comparing the amount of a THAP-family polypeptide or of a THAP-family RNA species encoding a THAP-family polypeptide within said biological sample with a level detected in or expected from a control sample. An increased amount of said THAP-family polypeptide or said THAP-family RNA species within said biological sample compared to said level detected in or expected from said control sample indicates that said mammal has an elevated level of THAP-family expression, and a decreased amount of said THAP-family polypeptide or said THAP-family RNA species within said biological sample compared to said level detected in or expected from said control sample indicates that said mammal has a reduced level of expression of a THAP-family member.

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining THAP-family protein and/or nucleic acid expression as well as THAP-family activity, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant THAP-family expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with a THAP-family protein, nucleic acid expression or activity. For example, mutations in a THAP-family gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby phophylactically treat an individual prior to

the onset of a disorder characterized by or associated with a THAP-family protein, nucleic acid expression or activity.

Accordingly, the methods of the present invention are applicable generally to diseases related to regulation of apoptosis, including but not limited to disorders characterized by unwanted cell proliferation or generally aberrant control of differentiation, for example neoplastic or hyperplastic disorders, as well as disorders related to proliferation or lack thereof of endothelial cells, inflammatory disorders and neurodegenerative disorders.

Diagnostic Assays

An exemplary method for detecting the presence (quantitative or not) or absence of a THAP-family protein or nucleic acid in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting a THAP-family protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes THAP-family protein such that the presence of the THAP-family protein or nucleic acid is detected in the biological sample. A preferred agent for detecting a THAP-family mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to a THAP-family mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length THAP-family nucleic acid, such as the nucleic acid of SEQ ID NO: 160 such as a nucleic acid of at least 15, 30, 50, 100, 250, 400, 500 or 1000 nucleotides in length and sufficient to specifically hybridize under stringent conditions to a THAP-family mRNA or genomic DNA or a portion of a THAP-family nucleic acid. Other suitable probes for use in the diagnostic assays of the invention are described herein.

In preferred embodiments, the subject method can be characterized by generally comprising detecting, in a tissue sample of the subject (e.g. a human patient), the presence or absence of a genetic lesion characterized by at least one of (i) a mutation of a gene encoding one of the subject THAP-family proteins or (ii) the mis-expression of a THAP-family gene. To illustrate, such genetic lesions can be detected by ascertaining the existence of at least one of (i) a deletion of one or more nucleotides from a THAP-family gene, (ii) an addition of one or more nucleotides to such a THAP-family gene, (iii) a substitution of one or more nucleotides of a THAP-family gene, (iv) a gross

chromosomal rearrangement or amplification of a THAP-family gene, (v) a gross alteration in the level of a messenger RNA transcript of a THAP-family gene, (vi) aberrant modification of a THAP-family gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a THAP-family gene, and (viii) a non-wild type level of a THAP-family -target protein.

A preferred agent for detecting a THAP-family protein is an antibody capable of binding to a THAP-family protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect a THAP-family mRNA, protein, or genomic DNA in a biological sample in vitro as well as in vivo. For example, in vitro techniques for detection of a THAP-family mRNA include Northern hybridizations and in situ hybridizations. In vitro techniques for detection of a THAP-family protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. In vitro techniques for detection of a THAP-family genomic DNA include Southern hybridizations. Furthermore, in vivo techniques for detection of a THAP-family protein include introducing into a subject a labeled anti-THAP-family antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In yet another exemplary embodiment, aberrant methylation patterns of a THAP-family gene can be detected by digesting genomic DNA from a patient sample with one or more restriction endonucleases that are sensitive to methylation and for

which recognition sites exist in the THAP-family gene (including in the flanking and intronic sequences). See, for example, Buiting et al. (1994) Human Mol Genet 3:893-895. Digested DNA is separated by gel electrophoresis, and hybridized with probes derived from, for example, genomic or cDNA sequences. The methylation status of the THAP-family gene can be determined by comparison of the restriction pattern generated from the sample DNA with that for a standard of known methylation.

Furthermore, gene constructs such as those described herein can be utilized in diagnostic assays to determine if a cell's growth or differentiation state is no longer dependent on the regulatory function of a THAP-family protein, e.g. in determining the phenotype of a transformed cell. Such knowledge can have both prognostic and therapeutic benefits. To illustrate, a sample of cells from the tissue can be obtained from a patient and dispersed in appropriate cell culture media, a portion of the cells in the sample can be caused to express a recombinant THAP-family protein or a THAP-family target protein, e.g. by transfection with a expression vector described herein, or to increase the expression or activity of an endogenous THAP-family protein or THAP-family target protein, and subsequent growth of the cells assessed. The absence of a change in phenotype of the cells despite expression of the THAP-family or THAP-family target protein may be indicative of a lack of dependence on cell regulatory pathways which includes the THAP-family or THAP-family target protein, e.g. THAP-family- or THAP-family target-mediated transcription. Depending on the nature of the tissue of interest, the sample can be in the form of cells isolated from, for example, a blood sample, an exfoliated cell sample, a fine needle aspirant sample, or a biopsied tissue sample. Where the initial sample is a solid mass, the tissue sample can be minced or otherwise dispersed so that cells can be cultured, as is known in the art.

In yet another embodiment, a diagnostic assay is provided which detects the ability of a THAP-family gene product, e.g., isolated from a biopsied cell, to bind to other cellular proteins. For instance, it will be desirable to detect THAP-family mutants which, while expressed at appreciable levels in the cell, are defective at binding a THAP-family target protein (having either diminished or enhanced binding affinity). Such mutants may arise, for example, from mutations, e.g., point mutants, which may be impractical to detect by the diagnostic DNA sequencing techniques or by the immunoassays described above. The present invention accordingly further contemplates

diagnostic screening assays which generally comprise cloning one or more THAP-family genes from the sample cells, and expressing the cloned genes under conditions which permit detection of an interaction between that recombinant gene product and a target protein, e.g., for example the THAP1 gene and a target PAR4 protein or a PML-NB protein. As will be apparent from the description of the various drug screening assays set forth below, a wide variety of techniques can be used to determine the ability of a THAP-family protein to bind to other cellular components. These techniques can be used to detect mutations in a THAP-family gene which give rise to mutant proteins with a higher or lower binding affinity for a THAP-family target protein relative to the wild-type THAP-family. Conversely, by switching which of the THAP-family target protein and THAP-family protein is the "bait" and which is derived from the patient sample, the subject assay can also be used to detect THAP-family target protein mutants which have a higher or lower binding affinity for a THAP-family protein relative to a wild type form of that THAP-family target protein.

In an exemplary embodiment, a PAR4 or a PMB-NB protein (e.g. wild-type) can be provided as an immobilized protein (a "target"), such as by use of GST fusion proteins and glutathione treated microtitre plates. A THAP1 gene (a "sample" gene) is amplified from cells of a patient sample, e.g., by PCR, ligated into an expression vector, and transformed into an appropriate host cell. The recombinantly produced THAP1 protein is then contacted with the immobilized PAR4 or PMB-NB protein, e.g., as a lysate or a semi-purified preparation, the complex washed, and the amount of PAR4 or PMB-NB protein /THAP1 complex determined and compared to a level of wild-type complex formed in a control. Detection can be by, for instance, an immunoassay using antibodies against the wild-type form of the THAP1 protein, or by virtue of a label provided by cloning the sample THAP1 gene into a vector which provides the protein as a fusion protein including a detectable tag. For example, a myc epitope can be provided as part of a fusion protein with the sample THAP1 gene. Such fusion proteins can, in addition to providing a detectable label, also permit purification of the sample THAP1 protein from the lysate prior to application to the immobilized target. In yet another embodiment of the subject screening assay, the two hybrid assay, described in the appended examples, can be used to detect mutations in either a THAP-family gene or THAP-family target gene which alter complex formation between those two proteins.

Accordingly, the present invention provides a convenient method for detecting mutants of THAP-family genes encoding proteins which are unable to physically interact with a THAP-family target "bait" protein, which method relies on detecting the reconstitution of a transcriptional activator in a THAP-family/THAP-family target-dependent fashion.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a serum sample isolated by conventional means from a subject. In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting a THAP-family protein, mRNA, or genomic DNA, such that the presence of a THAP-family protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of a THAP-family protein, mRNA or genomic DNA in the control sample with the presence of a THAP-family protein, mRNA or genomic DNA in the test sample. The invention also encompasses kits for detecting the presence of THAP-family protein, mRNA or genomic DNA in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting a THAP-family protein or mRNA or genomic DNA in a biological sample; means for determining the amount of a THAP-family member in the sample; and means for comparing the amount of THAP-family member in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect THAP-family protein or nucleic acid.

In certain embodiments, detection involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Pat. Nos. 4,683,195 and 4,683,202, the disclosures of which are incorporated herein by reference in their entireties), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegren et al. (1988) Science 241:1077-1080; and Nakazawa et al. (1994) PNAS 91:360-364, the disclosures of which are incorporated herein by reference in their entireties), the latter of which can be particularly useful for detecting point mutations in the THAP-family-gene (see Abravaya et al. (1995) Nucleic Acids Res. 23:675-682, the disclosure of which is incorporated herein by reference in its entirety). This method can

include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a THAP-family gene under conditions such that hybridization and amplification of the THAP-family-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Genotyping assays for diagnostics generally require the previous amplification of the DNA region carrying the biallelic marker of interest. However, ultrasensitive detection methods which do not require amplification are also available. Methods well-known to those skilled in the art that can be used to detect biallelic polymorphisms include methods such as, conventional dot blot analyzes, single strand conformational polymorphism analysis (SSCP) described by Orita et al., PNAS 86 : 2766-2770 (1989), the disclosure of which is incorporated herein by reference in its entirety, denaturing gradient gel electrophoresis (DGGE), heteroduplex analysis, mismatch cleavage detection, and other conventional techniques as described in Sheffield et al. (1991), White et al. (1992), and Grompe et al. (1989 and 1993) (Sheffield, V.C. et al, Proc. Natl. Acad. Sci. U.S.A 49:699-706 (1991); White, M.B. et al., Genomics 12:301-306 (1992); Grompe, M. et al., Proc. Natl. Acad. Sci. U.S.A 86:5855-5892 (1989); and Grompe, M. Nature Genetics 5:111-117 (1993), the disclosures of which are incorporated herein by reference in their entireties). Another method for determining the identity of the nucleotide present at a particular polymorphic site employs a specialized exonuclease-resistant nucleotide derivative as described in U.S. patent 4,656,127, the disclosure of which is incorporated herein by reference in its entirety. Further methods are described as follows.

The nucleotide present at a polymorphic site can be determined by sequencing methods. In a preferred embodiment, DNA samples are subjected to PCR amplification before sequencing as described above. DNA sequencing methods are described in "Sequencing Of Amplified Genomic DNA And Identification Of Single Nucleotide Polymorphisms". Preferably, the amplified DNA is subjected to automated dideoxy

terminator sequencing reactions using a dye-primer cycle sequencing protocol. Sequence analysis allows the identification of the base present at the biallelic marker site.

In microsequencing methods, the nucleotide at a polymorphic site in a target DNA is detected by a single nucleotide primer extension reaction. This method involves appropriate microsequencing primers which, hybridize just upstream of the polymorphic base of interest in the target nucleic acid. A polymerase is used to specifically extend the 3' end of the primer with one single ddNTP (chain terminator) complementary to the nucleotide at the polymorphic site. Next the identity of the incorporated nucleotide is determined in any suitable way. Typically, microsequencing reactions are carried out using fluorescent ddNTPs and the extended microsequencing primers are analyzed by electrophoresis on ABI 377 sequencing machines to determine the identity of the incorporated nucleotide as described in EP 412 883, the disclosure of which is incorporated herein by reference in its entirety. Alternatively capillary electrophoresis can be used in order to process a higher number of assays simultaneously. Different approaches can be used for the labeling and detection of ddNTPs. A homogeneous phase detection method based on fluorescence resonance energy transfer has been described by Chen and Kwok (1997) and,Chen and Kwok (*Nucleic Acids Research* 25:347-353 1997) and Chen et al. (*Proc. Natl. Acad. Sci. USA* 94/20 10756-10761,1997), the disclosures of which are incorporated herein by reference in their entireties). In this method, amplified genomic DNA fragments containing polymorphic sites are incubated with a 5'-fluorescein–labeled primer in the presence of allelic dye-labeled dideoxyribonucleoside triphosphates and a modified Taq polymerase. The dye-labeled primer is extended one base by the dye-terminator specific for the allele present on the template. At the end of the genotyping reaction, the fluorescence intensities of the two dyes in the reaction mixture are analyzed directly without separation or purification. All these steps can be performed in the same tube and the fluorescence changes can be monitored in real time. Alternatively, the extended primer may be analyzed by MALDI-TOF Mass Spectrometry. The base at the polymorphic site is identified by the mass added onto the microsequencing primer (see Haff and Smirnov, 1997, Genome Research, 7:378-388, 1997, the disclosure of which is incorporated herein by reference in its entirety). In another example, Pastinen et al.,

(Genome Research 7:606-614, 1997), the disclosure of which is incorporated herein by reference in its entirety) describe a method for multiplex detection of single nucleotide polymorphism in which the solid phase minisequencing principle is applied to an oligonucleotide array format. High-density arrays of DNA probes attached to a solid support (DNA chips) are further described below.

Other assays include mismatch detection assays, based on the specificity of polymerases and ligases. Polymerization reactions places particularly stringent requirements on correct base pairing of the 3' end of the amplification primer and the joining of two oligonucleotides hybridized to a target DNA sequence is quite sensitive to mismatches close to the ligation site, especially at the 3' end.

A preferred method of determining the identity of the nucleotide present at an allele involves nucleic acid hybridization. Any hybridization assay may be used including Southern hybridization, Northern hybridization, dot blot hybridization and solid-phase hybridization (see Sambrook et al., Molecular Cloning - A Laboratory Manual, Second Edition, Cold Spring Harbor Press, N.Y., 1989), the disclosure of which is incorporated herein by reference in its entirety). Hybridization refers to the formation of a duplex structure by two single stranded nucleic acids due to complementary base pairing. Hybridization can occur between exactly complementary nucleic acid strands or between nucleic acid strands that contain minor regions of mismatch. Specific probes can be designed that hybridize to one form of a biallelic marker and not to the other and therefore are able to discriminate between different allelic forms. Allele-specific probes are often used in pairs, one member of a pair showing perfect match to a target sequence containing the original allele and the other showing a perfect match to the target sequence containing the alternative allele. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Stringent, sequence specific hybridization conditions, under which a probe will hybridize only to the exactly complementary target sequence are well known in the art (Sambrook et al., 1989). The detection of hybrid duplexes can be carried out by a number of methods. Various detection assay formats are well known which utilize detectable labels bound to either the target or the probe to enable detection of the hybrid duplexes. Typically,

hybridization duplexes are separated from unhybridized nucleic acids and the labels bound to the duplexes are then detected. Further, standard heterogeneous assay formats are suitable for detecting the hybrids using the labels present on the primers and probes. (see Landegren U. et al., Genome Research, 8:769-776,1998, the disclosure of which is incorporated herein by reference in its entirety).

Hybridization assays based on oligonucleotide arrays rely on the differences in hybridization stability of short oligonucleotides to perfectly matched and mismatched target sequence variants. Efficient access to polymorphism information is obtained through a basic structure comprising high-density arrays of oligonucleotide probes attached to a solid support (e.g., the chip) at selected positions. Chips of various formats for use in detecting biallelic polymorphisms can be produced on a customized basis by Affymetrix (GeneChip), Hyseq (HyChip and HyGnostics), and Protogene Laboratories.

In general, these methods employ arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from an individual which, target sequences include a polymorphic marker. EP 785280, the disclosure of which is incorporated herein by reference in its entirety, describes a tiling strategy for the detection of single nucleotide polymorphisms. Briefly, arrays may generally be "tiled" for a large number of specific polymorphisms, further described in PCT application No. WO 95/11995, the disclosure of which is incorporated herein by reference in its entirety. Upon completion of hybridization with the target sequence and washing of the array, the array is scanned to determine the position on the array to which the target sequence hybridizes. The hybridization data from the scanned array is then analyzed to identify which allele or alleles of the biallelic marker are present in the sample. Hybridization and scanning may be carried out as described in PCT application No. WO 92/10092 and WO 95/11995 and US patent No. 5,424,186, the disclosures of which are incorporated herein by reference in their entireties. Solid supports and polynucleotides of the present invention attached to solid supports are further described in "Oligonucleotide Probes And Primers".

DETECTING CHEMOKINES

Some aspects of the present invention relate to the detection of chemokines by contacting a chemokine or a sample containing a chemokine with a THAP-type

chemokine-binding agent. In some embodiments, the chemokines or the THAP-type chemokine-binding agents are labeled. Many labels and methods of conjugating such labels to a chemokine or a THAP-type chemokine-binding agent are known in the art. Additionally, labeled molecules, such as antibodies, which have an affinity for a THAP-type chemokine-binding agent can be used to detect the chemokine that is bound to a THAP-type chemokine-binding agent using a number of assay formats that are well known in the art.

An exemplary method for detecting the presence (quantitative or not) or absence of a chemokine, including, but not limited to, a chemokine in a biological sample, involves obtaining a chemokine or a sample containing a chemokine and contacting it with a compound or an agent capable of detecting the chemokine. In some embodiments, such an agent is a THAP-type chemokine-binding agent. Chemokines which can be detected using a method that employs a THAP-type chemokine-binding agent include, but are not limited to, XCL1, XCL2, CCL1, CCL2, CCL3, CCL3L1, SCYA3L2, CCL4, CCL4L, CCL5, CCL6, CCL7, CCL8, SCYA9, SCYA10, CCL11, SCYA12, CCL13, CCL14, CCL15, CCL16, CCL17, CCL18, CCL19, CCL20, CCL21, CCL22, CCL23, CCL24, CCL25, CCL26, CCL27, CCL28, clone 391, CARP CC-1, CCL1, CK-1, regakine-1, K203, CXCL1, CXCL1P, CXCL2, CXCL3, PF4, PF4V1, CXCL5, CXCL6, PPBP, SPBPBP, IL8, CXCL9, CXCL10, CXCL11, CXCL12, CXCL14, CXCL15, CXCL16, NAP-4, LFCA-1, Scyba, JSC, VHSV-induced protein, CX3CL1 and fCL1.

In some embodiments, the detection method comprises detecting, in a biological sample, such as a tissue or fluid sample from a subject (such as, a human patient), the presence or absence of a chemokine by contacting the biological sample with a THAP-type chemokine-binding agent and detecting a complex between the chemokine and the THAP-type chemokine-binding agent or detecting a THAP-type chemokine-binding agent which was previously bound to the chemokine but which has been released from the chemokine.

In some embodiments of the present invention, the THAP-type chemokine-binding agent is labeled directly. In other embodiments, the THAP-type chemokine-binding agent is detected using a labeled antibody having affinity for the THAP-type chemokine-binding agent. Such antibodies may directly carry the detectable label or be

recognized by a labeled second antibody. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the antibody or other detectable molecule, is intended to encompass direct labeling of the antibody or molecule by coupling (i.e., physically linking) a detectable substance to the antibody or molecule, as well as indirect labeling of the antibody or molecule by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a THAP-type chemokine-binding agent with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. Accordingly, the detection method can be used to detect a chemokine in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of a chemokine include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vivo* techniques for detection of a chemokine include introducing into a subject a labeled THAP-type chemokine-binding agent. For example, the THAP-type chemokine-binding agent can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

Other aspects of the present invention relate to a system for chemokine detection. Such a chemokine detection system comprises a THAP-type chemokine-binding agent bound to a solid support. A number of adequate solid support materials are known in the art and include, but are not limited to, cellulose, nylon or other polymer backings, plastics such as microtiter plates, synthetic beads and resins such as sepharose, glass, magnetic beads, latex particles, sheep (or other animal) red blood cells, duracytes and others. Suitable methods for immobilizing the THAP-type chemokine-binding agent to the solid support are well known in the art.

Some embodiments of the present invention relate to kits which comprise a THAP-type chemokine-binding agent and instructions which describe detecting or inhibiting chemokines with the THAP-type chemokine-binding agent. For example, the kit includes an ampule of THAP-type chemokine-binding agent that is stored so as to prevent damage or inactivation of the agent upon prolonged storage. Such methods can

include, but are not limited to, lyophilization and freezing in an appropriate buffer. The kit also can contain chemokines to serve as a positive control sample when the kit is used for chemokine binding, detection or inhibition.

In some embodiments of the present invention, kits are packaged containing a heterogeneous mixture of THAP-type chemokine-binding agents, wherein each of the agents has a different affinity for one or more chemokines. Alternatively, some kits comprise a panel of THAP-type chemokine-binding agents, wherein each THAP-type chemokine binding agent has a different affinity for a particular chemokine. For example, the kit can comprise a panel of three THAP-type chemokine-binding agents, wherein the first agent has a high affinity for SLC but a low affinity for CXCL9, the second agent has a moderate affinity for both SLC and CXCL9, and the third agent has a low affinity for SLC and a high affinity for CXCL9. Panels of THAP-type chemokine-binding agents can be larger or small than that exemplified above and the number and types of chemokines that are detected can be more or less than that exemplified above. Kits containing such panels of THAP-type chemokine-binding agents can be used to reliably distinguish mixed samples of chemokines. Additionally, such panels can be used to bind or inhibit multiple different chemokines in a mixed chemokine sample. Having generally described this invention, a further understanding can be obtained by reference to certain specific examples which are provided herein for purposes of illustration only, and are not intended to be limiting unless otherwise specified.

EXAMPLES

EXAMPLE 1

Isolation of the THAP1 cDNA in a two-hybrid screen with chemokine SLC/CCL21

In an effort to define the function of novel HEVEC proteins and the cellular pathways involved, we used different baits to screen a two-hybrid cDNA library generated from microvascular human HEV endothelial cells (HEVEC). HEVEC were purified from human tonsils by immunomagnetic selection with monoclonal antibody MECA-79 as previously described (Girard and Springer (1995) Immunity 2:113-123). The SMART PCR cDNA library Construction Kit (Clontech, Palo Alto, CA, USA) was first used to generate full-length cDNAs from 1 μg HEVEC total RNA. Oligo-dT-primed HEVEC cDNA were then digested with SfiI and directionally cloned into pGAD424-Sfi, a two-hybrid vector generated by inserting a SfiI linker (5'-GAATTCGGCCATTATGGCCTGCAGGATCCGGCCGCCTCGGCCCAGGATCC-3') (SEQ ID NO: 181) between EcoRI and BamHI cloning sites of pGAD424 (Clontech). The resulting pGAD424-HEVEC cDNA two-hybrid library (mean insert size > 1 kb, ~ $3x10^6$ independant clones) was amplified in *E. coli*. To identify potential protein partners of chemokine SLC/6Ckine, screening of the two-hybrid HEVEC cDNA library was performed using as bait a cDNA encoding the mature form of human SLC/CCL21 (amino acids 24-134, GenBank Accession No: NP_002980, SEQ ID NO: 182), amplified by PCR from HEVEC RNA with primers hSLC.5' (5'-GCGGGATCCGTAGTGATGGAGGGGCTCAGGACTGTTG-3') (SEQ ID NO: 183) and hSLC.3' (5'-GCGGGATCCCTATGGCCCTTTAGGGGTCTGTGACC-3') (SEQ ID NO: 184), digested with BamHI and inserted into the BamHI cloning site of MATCHMAKER two-hybrid system 2 vector pGBT9 (Clontech). Briefly, pGBT9-SLC was cotransformed with the pGAD424-HEVEC cDNA library in yeast strain Y190 (Clontech). $1.5x10^7$ yeast transformants were screened and positive protein interactions were selected by His auxotrophy. The plates were incubated at 30°C for 5 days. Plasmid DNA was extracted from positive colonies and used to verify the specificity of the interaction by cotransformation in AH109 with pGBT9-SLC or control baits pGBT9, pGBT9-lamin. Eight independent clones isolated in this two-hybrid screen were characterized. They were found to correspond to a unique human cDNA encoding a

novel human protein of 213 amino acids, designated THAP1, that exhibits 93% identity with its mouse orthologue (Figure 1A). The only noticeable motifs in the THAP1 predicted protein sequence were a short proline-rich domain in the middle part and a consensus nuclear localization sequence (NLS) in the carboxy terminal part (Figure 1B). Databases searches with the THAP1 sequence failed to reveal any significant similarity to previously characterized proteins with the exception of the first 90 amino acids that may define a novel protein motif associated with apoptosis, hereafter referred to as THAP domain (see Figure 1B, Figures 9A-9C, and Figure 10).

EXAMPLE 2

Northern Blot

To determine the tissue distribution of THAP1 mRNA, we performed Northern blot analysis of 12 different adult human tissues (Fig 2). Multiple Human Tissues Northern Blots (CLONTECH) were hydridized according to manufacturer's instructions. The probe was a PCR product corresponding to the THAP1 ORF, ^{32}P-labeled with the Prime-a-Gene Labeling System (PROMEGA).A 2.2-kb mRNA band was detected in brain, heart, skeletal muscle, kidney, liver, and placenta. In addition to the major 2.2 kb band, lower molecular weight bands were detected, that are likely to correspond to alternative splicing or polyadenylation of the THAP1 pre-mRNA. The presence of THAP1 mRNAs in many different tissues suggests that THAP1 has a widespread, although not ubiquitous, tissue distribution in the human body.

EXAMPLE 3

Analysis of the subcellular THAP1 localization

To analyze the subcellular localization of the THAP1 protein, the THAP1 cDNA was fused to the coding sequence of GFP (Green Fluorescent Protein). The full-length coding region of THAP1 was amplified by PCR from HEVEC cDNA with primers *2HMR10* (5'-CCGAATTCAGGATGGTGCAGTCCTGCTCCGCCT-3') (SEQ ID NO: 185) and *2HMR9* (5'-CGCGGATCCTGCTGGTACTTCAACTATTTCAAAGTAGTC-3') (SEQ ID NO: 186), digested with EcoRI and BamHI, and cloned in frame downstream of the Enhanced Green Fluorescent Protein (EGFP) ORF in pEGFP.C2 vector (Clontech) to generate pEGFP.C2-THAP1. The GFP/THAP1 expression

construct was then transfected into human primary endothelial cells from umbilical vein (HUVEC, PromoCell, Heidelberg, Germany). HUVEC were grown in complete ECGM medium (PromoCell, Heidelberg, Germany), plated on coverslips and transiently transfected in RPMI medium using GeneJammer transfection reagent according to manufacturer instructions (Stratagene, La Jolla, CA, USA). Analysis by fluorescence microscopy 24h later revealed that the GFP/THAP1 fusion protein localizes exclusively in the nucleus with both a diffuse distribution and an accumulation into speckles while GFP alone exhibits only a diffuse staining over the entire cell. To investigate the identity of the speckled domains with which GFP/THAP1 associates, we used indirect immunofluorescence microscopy to examine a possible colocalization of the nuclear dots containing GFP/THAP1 with known nuclear domains (replication factories, splicing centers, nuclear bodies).

Cells transfected with GFP-tagged expression constructs were allowed to grow for 24 h to 48 h on coverslips. Cells were washed twice with PBS, fixed for 15 min at room temperature in PBS containing 3.7% formaldehyde, and washed again with PBS prior to neutralization with 50mM NH_4Cl in PBS for 5 min at room temperature. Following one more PBS wash, cells were permeabilized 5 min at room temperature in PBS containing 0.1% Triton-X100, and washed again with PBS. Permeabilized cells were then blocked with PBS-BSA (PBS with 1% bovine serum albumin) for 10' and then incubated 2 hr at room temperature with the following primary antibodies diluted in PBS-BSA: rabbit polyclonal antibodies against human Daxx (1/50, M-112, Santa Cruz Biotechnology) or mouse monoclonal antibodies anti-PML (mouse IgG1, 1/30, mAb PG-M3 from Dako, Glostrup, Denmark). Cells were then washed three times 5 min at room temperature in PBS-BSA, and incubated for 1 hr with Cy3 (red fluorescence)-conjugated goat anti-mouse or anti-rabbit IgG (1/1000, Amersham Pharmacia Biotech) secondary antibodies, diluted in PBS-BSA. After extensive washing in PBS, samples were air dried and mounted in Mowiol. Images were collected on a Leica confocal laser scanning microscope. The GFP (green) and Cy3 (red) fluorescence signals were recorded sequentially for identical image fields to avoid cross-talk between the channels.

This analysis revealed that GFP-THAP1 staining exhibits a complete overlap with the staining pattern obtained with antibodies directed against PML. The

colocalization of GFP/THAP1 and PML was observed both in nuclei with few PML-NBs (less than ten) and in nuclei with a large number of PML-NBs. Indirect immunofluorescence staining with antibodies directed against Daxx, another well characterized component of PML-NBs, was performed to confirm the association of GFP/THAP1 with PML-NBs. We found a complete colocalization of GFP/THAP1 and Daxx in PML-NBs. Together, these results reveal that THAP1 is a novel protein associated with PML-NBs.

EXAMPLE 4

Identification of proteins interacting with THAP1 in human HEVECs: two-hybrid assay

THAP1 forms a complex with the pro-apoptotic protein PAR4

To identify potential protein partners of THAP1, screening of the two-hybrid HEVEC cDNA library was performed using as a bait the human THAP1 full length cDNA inserted into the MATCHMAKER two-hybrid system 3 vector pGBKT7 (Clontech). Briefly, the full-length coding region of THAP1 was amplified by PCR from HEVEC cDNA with primers *2HMR10* (5'-CCGAATTCAGGATGGTGCAGTCCTGCTCCGCCT-3') (SEQ ID NO: 187) and *2HMR9* (5'-CGCGGATCCTGCTGGTACTTCAACTATTTCAAAGTAGTC-3') (SEQ ID NO: 188), digested with EcoRI and BamHI, and cloned in frame downstream of the Gal4 Binding Domain (Gal4-BD) in pGBKT7 vector to generate pGBKT7-THAP1. pGBKT7-THAP1 was then cotransformed with the pGAD424-HEVEC cDNA library in yeast strain AH109 (Clontech). $1.5x10^7$ yeast transformants were screened and positive protein interactions were selected by His and Ade double auxotrophy according to manufacturer's instructions (MATCHMAKER two-hybrid system 3, Clontech). The plates were incubated at 30°C for 5 days. Plasmid DNA was extracted from these positive colonies and used to verify the specificity of the interaction by cotransformation in AH109 with pGBKT7-THAP1 or control baits pGBKT7, pGBKT7-lamin and pGBKT7-hevin. Three clones which specifically interacted with THAP1 were obtained in the screen ; sequencing of these clones revealed three identical library plasmids that corresponded to a partial cDNA coding for the last 147 amino acids (positions 193-342) of the human pro-apoptotic protein PAR4 (Fig 3A). Positive interaction between THAP1 and Par4 was confirmed using full length Par4 bait

(pGBKT-Par4) and prey (pGADT7-Par4). Full-length human Par4 was amplified by PCR from human thymus cDNA (Clontech), with primers *Par4.8* (5'-GCGGAATTCATGGCGACCGGTGGCTACCGGACC-3') (SEQ ID NO: 189) and *Par4.5* (5'-GCGGGATCCCTCTACCTGGTCAGCTGACCCACAAC-3') (SEQ ID NO: 190), digested with EcoRI and BamHI, and cloned in pGBKT7 and pGADT7 vectors, to generate pGBKT7-Par4 and pGADT7-Par4. Positive interaction between THAP1 and Par4 was confirmed by cotransformation of AH109 with pGBKT7-THAP1 and pGADT7-Par4 or pGBKT7-Par4 and pGADT7-THAP1 and selection of transformants by His and Ade double auxotrophy according to manufacturer's instructions (MATCHMAKER two-hybrid system 3, Clontech). To generate pGADT7-THAP1, the full-length coding region of THAP1 was amplified by PCR from HEVEC cDNA with primers *2HMR10* (5'-CCGAATTCAGGATGGTGCAGTCCTGCTCCGCCT-3') (SEQ ID NO: 191) and *2HMR9* (5'-CGCGGATCCTGCTGGTACTTCAACTATTTCAAAGTAGTC-3') (SEQ ID NO: 192), digested with EcoRI and BamHI, and cloned in frame downstream of the Gal-4 Activation Domain (Gal4-AD) in pGADT7 two-hybrid vector (Clontech).

We then examined whether the leucine zipper/death domain at the C-terminus of Par4, previously shown to be involved in Par4 binding to WT-1 and aPKC, was required for the interaction between THAP1 and Par4. Two Par4 mutants were constructed for that purpose, Par4Δ and Par4DD. Par4Δ lacks the leucine zipper/death domain while Par4DD contains this domain. pGBKT7-Par4Δ(amino acids 1-276) and pGADT7-Par4Δ were constructed by sub-cloning a EcoRI-BglII fragment from pGADT7-Par4 into the EcoRI and BamHI sites of pGBKT7 and pGADT7. Par4DD (amino acids 250-342) was amplified by PCR, using pGBKT7-Par4 as template, with primers *Par4.4* (5'-CGCGAATTCGCCATCATGGGGTTCCCTAGATATAACAGGGATGCAA-3') (SEQ ID NO: 193) and *Par4.5*, and cloned into the EcoRI and BamHI sites of pGBKT7 and pGADT7 to obtain pGBKT7-Par4DD and pGADT7-Par4DD. Two-hybrid interaction between THAP1 and Par4 mutants was tested by cotransformation of AH109 with pGBKT7-THAP1 and pGADT7-Par4Δ or pGADT7-Par4DD and selection of transformants by His and Ade double auxotrophy according to manufacturer's instructions (MATCHMAKER two-hybrid system 3, Clontech). We found that the

Par4 leucine zipper/death domain (Par4DD) is not only required but also sufficient for the interaction with THAP1 (Fig 3A). Similar results were obtained when two-hybrid experiments were performed in the opposite orientation using Par4 or Par4 mutants (Par4Δ and Par4DD) as baits instead of THAP1 (Fig 3A).

5

EXAMPLE 5

In vitro THAP1/Par4 interaction assay

To confirm the interaction observed in yeast, we performed *in vitro* GST pull down assays. Par4DD, expressed as a GST-tagged fusion protein and immobilized on 10 glutathione sepharose, was incubated with radiolabeled *in vitro* translated THAP1. To generate the GST-Par4DD expression vector, Par4DD (amino acids 250-342) was amplified by PCR with primers *Par4.10* (5'-GCCGGATCCGGGTTCCCTAGATATAACAGGGATGCAA-3') (SEQ ID NO: 194) and *Par4.5*, and cloned in frame downstream of the Glutathion S-Transferase ORF, into 15 the BamHI site of the pGEX-2T prokaryotic expression vector (Amersham Pharmacia Biotech, Saclay, France). GST-Par4DD(amino acids 250-342) fusion protein encoded by plasmid pGEX-2T-Par4DD and control GST protein encoded by plasmid pGEX-2T, were then expressed in *E. coli* DH5α and purified by affinity chromatography with glutathione sepharose according to supplier's instructions (Amersham Pharmacia 20 Biotech). The yield of proteins used in GST pull-down assays was determined by SDS-Polyarylamide Gel Electrophoresis (PAGE) and Coomassie blue staining analysis. *In vitro*-translated THAP1 was generated with the TNT-coupled reticulocyte lysate system (Promega, Madison, WI, USA) using pGBKT7-THAP1 vector as template. 25 µl of ^{35}S-labeled wild-type THAP1 was incubated with immobilized GST-Par4 or GST 25 proteins overnight at 4 °C, in the following binding buffer : 10 mM NaPO4 pH 8.0, 140 mM NaCl, 3 mM MgCl2, 1mM dithiothreitol (DTT), 0.05% NP40, and 0.2 mM phenylmethyl sulphonyl fluoride (PMSF), 1 mM Na vanadate, 50mM β Glycerophosphate, 25 µg/ml chymotrypsine, 5 µg/ml aprotinin, and 10 µg/ml leupeptin. Beads were then washed 5 times in 1 ml binding buffer. Bound proteins were eluted 30 with 2X Laemmli SDS-PAGE sample buffer, fractionated by 10% SDS–PAGE and visualized by fluorography using Amplify (Amersham Pharmacia Biotech). As

expected, GST/Par4DD interacted with THAP1 (Fig 3B). In contrast, THAP1 failed to interact with GST beads.

EXAMPLE 6

In vivo THAP1/Par4 interaction assay

To provide further evidence for a physiological interaction between THAP1 and Par4 *in vivo* interactions between THAP1 and PAR4 were investigated. For that purpose, confocal immunofluorescence microscopy was used to analyze the subcellular localization of epitope-tagged Par4DD in primary human endothelial cells transiently cotransfected with pEF-*myc*Par4DD eukaryotic expression vector and GFP or GFP-THAP1 expression vectors (pEGFP.C2 and pEGFP.C2-THAP1, respectively). To generate pEF-*myc*Par4DD, mycPar4DD (amino acids 250-342) was amplified by PCR using pGBKT7-Par4DD as template, with primers *myc.BD7* (5'-GCGCTCTAGAGCCATCATGGAGGAGCAGAAGCTGATC-3') (SEQ ID NO: 195) and *Par4.9* (5'-CTTGCGGCCGCCTCTACCTGGTCAGCTGACCCACAAC-3') (SEQ ID NO: 196), and cloned into the XbaI and NotI sites of the pEF-BOS expression vector (Mizushima and Nagata, Nucleic Acids Research, 18:5322, 1990). Primary human endothelial cells from umbilical vein (HUVEC, PromoCell, Heidelberg, Germany) were grown in complete ECGM medium (PromoCell, Heidelberg, Germany), plated on coverslips and transiently transfected in RPMI medium using GeneJammer transfection reagent according to manufacturer instructions (Stratagene, La Jolla, CA, USA). Cells co-transfected with pEF-*myc*Par4DD and GFP-tagged expression constructs were allowed to grow for 24 h to 48 h on coverslips. Cells were washed twice with PBS, fixed for 15 min at room temperature in PBS containing 3.7% formaldehyde, and washed again with PBS prior to neutralization with 50mM NH_4Cl in PBS for 5 min at room temperature. Following one more PBS wash, cells were permeabilized 5 min at room temperature in PBS containing 0.1% Triton-X100, and washed again with PBS. Permeabilized cells were then blocked with PBS-BSA (PBS with 1% bovine serum albumin) for 10' and then incubated 2 hr at room temperature with mouse monoclonal antibody anti-myc epitope (mouse IgG1, 1/200, Clontech) diluted in PBS-BSA. Cells were then washed three times 5 min at room temperature in PBS-BSA, and incubated for 1 hr with Cy3 (red fluorescence)-conjugated goat anti-mouse (1/1000, Amersham

Pharmacia Biotech) secondary antibodies, diluted in PBS-BSA. After extensive washing in PBS, samples were air dried and mounted in Mowiol. Images were collected on a Leica confocal laser scanning microscope. The GFP (green) and Cy3 (red) fluorescence signals were recorded sequentially for identical image fields to avoid cross-talk between the channels.

In cells transiently co-transfected with pEF-*myc*Par4DD and GFP expression vector, ectopically expressed myc-Par4DD was found to accumulate both in the cytoplasm and the nucleus of the majority of the cells. In contrast, transient cotransfection of pEF-*myc*Par4DD and GFP-THAP1 expression vectors dramatically shifted myc-Par4DD from a diffuse cytosolic and nuclear localization to a preferential association with PML-NBs. The effect of GFP-THAP1 on myc-Par4DD localization was specific since it was not observed with GFP-APS kinase-1 (APSK-1), a nuclear enzyme unrelated to THAP1 and apoptosis [Besset et al., Faseb J, 14:345-354, 2000]. This later result shows that GFP-THAP1 recruits myc-Par4DD at PML-NBs and provides *in vivo* evidence for a direct interaction of THAP1 with the pro-apoptotic protein Par4.

EXAMPLE 7

Identification of a novel arginine-rich Par4 binding motif

To identify the sequences mediating THAP1 binding to Par4, a series of THAP1 deletion constructs was generated. Both amino-terminal (THAP1-C1, -C2, -C3) and carboxy-terminal (THAP1-N1, -N2, -N3) deletion mutants (Figure 4A) were amplified by PCR using plasmid pEGFP.C2-THAP1 as a template and the following primers:
2HMR12 (5'-GCGGAATTCAAAGAAGATCTTCTGGAGCCACAGGAAC-3') (SEQ ID NO: 197)
and *2HMR9* (5'-CGCGGATCCTGCTGGTACTTCAACTATTTCAAAGTAGTC-3') (SEQ ID NO: 198) for THAP1-C1 (amino acids 90-213);
PAPM2 (5'-GCGGAATTCATGCCGCCTCTTCAGACCCCTGTTAA-3') (SEQ ID NO: 199)
and *2HMR9* for THAP1-C2 (amino acids 120-213);
PAPM3 (5'-GCGGAATTCATGCACCAGCGGAAAAGGATTCATCAG-3') (SEQ ID NO: 200)

and *2HMR9* for THAP1-C3 (amino acids 143-213);
2HMR10 (5'-CCGAATTCAGGATGGTGCAGTCCTGCTCCGCCT-3') (SEQ ID NO: 201)
and *2HMR17* (5'-GCGGGATCCCTTGTCATGTGGCTCAGTACAAAGAAATAT-3') (SEQ ID NO: 202) for THAP1-N1 (amino acids 1-90);
2HMR10 and *PAPN2* (5'-CGGGATCCTGTGCGGTCTTGAGCTTCTTTCTGAG-3') (SEQ ID NO: 203) for THAP1-N2 (amino acids 1-166); and
2HMR10 and *PAPN3* (5'-GCGGGATCCGTCGTCTTTCTCTTTCTGGAAGTGAAC-3') (SEQ ID NO: 204) for THAP1-N3 (amino acids 1-192).

The PCR fragments, thus obtained, were digested with EcoRI and BamHI, and cloned in frame downstream of the Gal4 Binding Domain (Gal4-BD) in pGBKT7 two-hybrid vector (Clontech) to generate pGBKT7-THAP1-C1, -C2, -C3, -N1, -N2 or -N3, or downstream of the Enhanced Green Fluorescent Protein (EGFP) ORF in pEGFP.C2 vector (Clontech) to generate pEGFP.C2-THAP1-C1, -C2, -C3, -N1, -N2 or -N3.

Two-hybrid interaction between THAP1 mutants and Par4DD was tested by cotransformation of AH109 with pGBKT7-THAP1-C1, -C2, -C3, -N1, -N2 or -N3 and pGADT7-Par4DD and selection of transformants by His and Ade double auxotrophy according to manufacturer's instructions (MATCHMAKER two-hybrid system 3, Clontech). Positive two-hybrid interaction with Par4DD was observed with mutants THAP1-C1, -C2, -C3, -and -N3 but not with mutants THAP1-N1 and -N2, suggesting the Par4 binding site is found between THAP1 residues 143 and 192.

THAP1 mutants were also tested in the *in vitro* THAP1/Par4 interaction assay. *In vitro*-translated THAP1 mutants were generated with the TNT-coupled reticulocyte lysate system (Promega, Madison, WI, USA) using pGBKT7-THAP1-C1, -C2, -C3, -N1, -N2 or -N3 vector as template. 25 µl of each ^{35}S-labelled THAP1 mutant was incubated with immobilized GST or GST-Par4 protein overnight at 4 °C, in the following binding buffer : 10 mM NaPO4 pH 8.0, 140 mM NaCl, 3 mM MgCl2, 1mM dithiothreitol (DTT), 0.05% NP40, and 0.2 mM phenylmethyl sulphonyl fluoride (PMSF), 1 mM Na Vanadate, 50mM β Glycerophosphate, 25 µg/ml chimotrypsine, 5 µg/ml aprotinin, 10 µg/ml Leupeptin. Beads were then washed 5 times in 1 ml binding buffer. Bound proteins were eluted with 2X Laemmli SDS-PAGE sample buffer, fractionated by 10% SDS–PAGE and visualized by fluorography using Amplify

(Amersham Pharmacia Biotech). As expected, THAP1-C1, -C2, -C3, -and -N3 interacted with GST/Par4DD (Figure 4B). In contrast, THAP1-N1 and -N2 failed to interact with GST/Par4DD beads.

Finally, Par4 binding activity of THAP1 mutants was also analyzed by the *in vivo* THAP1/Par4 interaction assay as described in Example 6 using pEF-*myc*Par4DD and pEGFP.C2-THAP1-C1, -C2, -C3, -N1, -N2 or -N3 expression vectors.

Essentially identical results were obtained with the three THAP1/Par4 interactions assays (Figure 4A). That is, the Par4 binding site was found between residues 143 and 192 of human THAP1. Comparison of this region with the Par4 binding domain of mouse ZIP kinase, another Par4-interacting protein, revealed the existence of a conserved arginine rich-sequence motif (SEQ ID NOs: 205, 263 and 15), that may correspond to the Par4 binding site (Figure 5A). Mutations in this arginine rich-sequence motif were generated by site directed mutagenesis. These two novel THAP1 mutants, THAP1 RR/AA (replacement of residues R171A and R172A) and THAP1ΔQRCRR (deletion of residues 168-172), were generated by two successive rounds of PCR using pEGFP.C2-THAP1 as template and primers *2HMR10* and *2HMR9* together with primers

RR/AA-1 (5'-CCGCACAGCAGCGATGCGCTGCTCAAGAACGGCAGCTTG-3') (SEQ ID NO: 206) and

RR/AA-2 (5'-CAAGCTGCCGTTCTTGAGCAGCGCATCGCTGCTGTGCGG-3') (SEQ ID NO: 207) for mutant THAP1 RR/AA or

primers *ΔRR-1* (5'-GCTCAAGACCGCACAGCAAGAACGGCAGCTTG-3'(SEQ ID NO: 208) and

ΔRR-2 (5'-CAAGCTGCCGTTCTTGCTGTGCGGTCTTGAGC-3') (SEQ ID NO: 209) for mutant THAP1ΔQRCRR. The resulting PCR fragments were digested with EcoRI and BamHI, and cloned in frame downstream of the Gal4 Binding Domain (Gal4-BD) in pGBKT7 two-hybrid vector (Clontech) to generate pGBKT7-THAP1-RR/AA and –Δ(QRCRR), or downstream of the Enhanced Green Fluorescent Protein (EGFP) ORF in pEGFP.C2 vector (Clontech) to generate pEGFP.C2-THAP1-RR/AA and –Δ(QRCRR). THAP1 RR/AA and THAP1ΔQRCRR THAP1 mutants were then tested in the three THAP1/Par4 interaction assays (two-hybrid assay, *in vitro* THAP1/Par4 interaction assay, *in vivo* THAP1/Par4 interaction assay) as described above for the THAP1-C1, -

C2, -C3, -N1, -N2 or -N3 mutants. This analysis revealed that the two mutants were deficient for interaction with Par4 in all three assays (Figure 5B), indicating that the novel arginine-rich sequence motif, we have identified, is a novel Par4 binding motif.

EXAMPLE 8

PAR4 is a novel component of PML-NBs that colocalizes with THAP1 *in vivo*

We then wished to determine if PAR4 colocalizes with THAP1 *in vivo* in order to provide further evidence for a physiological interaction between THAP1 and PAR4. We first analyzed Par4 subcellular localization in primary human endothelial cells. Confocal immunofluorescence microscopy using affinity-purified anti-PAR4 antibodies (Sells et al., 1997 ; Guo et al ; 1998) was performed on HUVEC endothelial cells fixed with methanol/acetone, which makes PML-NBs components accessible for antibodies (Sternsdorf et al., 1997). Cells were fixed in methanol for 5 min at –20°C, followed by incubation in cold acetone at –20°C for 30 sec. Permeabilized cells were then blocked with PBS-BSA (PBS with 1% bovine serum albumin) for 10' and then incubated 2 hr at room temperature with rabbit polyclonal antibodies against human Par4 (1/50, R-334, Santa Cruz Biotechnology, Santa Cruz, CA, USA) and mouse monoclonal antibody anti-PML (mouse IgG1, 1/30, mAb PG-M3 from Dako, Glostrup, Denmark). Cells were then washed three times 5 min at room temperature in PBS-BSA, and incubated for 1 hr with Cy3 (red fluorescence)-conjugated goat anti-rabbit IgG (1/1000, Amersham Pharmacia Biotech) and FITC-labeled goat anti-mouse-IgG (1/40, Zymed Laboratories Inc., San Francisco, CA, USA) secondary antibodies, diluted in PBS-BSA. After extensive washing in PBS, samples were air dried and mounted in Mowiol. Images were collected on a Leica confocal laser scanning microscope. The FITC (green) and Cy3 (red) fluorescence signals were recorded sequentially for identical image fields to avoid cross-talk between the channels. This analysis showed an association of PAR4 immunoreactivity with nuclear dot-like structures, in addition to diffuse nucleoplasmic and cytoplasmic staining. Double immunostaining with anti-PML antibodies, revealed that the PAR4 foci colocalize perfectly with PML-NBs in cell nuclei. Colocalization of Par4 with GFP-THAP1 in PML-NBs was analyzed in transfected HUVEC cells expressing ectopic GFP-THAP1. HUVEC were grown in complete ECGM medium (PromoCell, Heidelberg, Germany), plated on coverslips and

transiently transfected with GFP/THAP1 expression construct (pEGFP.C2-THAP1) in RPMI medium using GeneJammer transfection reagent according to manufacturer instructions (Stratagene, La Jolla, CA, USA). Analysis of transfected cells by indirect immunofluorescence microscopy 24h later, with anti-Par4 rabbit antibodies, revealed that all endogenous PAR4 foci colocalize with ectopic GFP-THAP1 in PML-NBs further confirming the association of the THAP1/PAR4 complex with PML-NBs *in vivo*.

EXAMPLE 9

PML recruits the THAP1/PAR4 complex to PML-NBs

Since it has been shown that PML plays a critical role in the assembly of PML-NBs by recruiting other components, we next wanted to determine whether PML plays a role in the recruitment of the THAP1/PAR4 complex to PML-NBs. For this purpose, we made use of the observation that both endogenous PAR4 and ectopic GFP-THAP1 do not accumulate in PML-NBs in human Hela cells. Expression vectors for GFP-THAP1 and HA-PML (or HA-SP100) were cotransfected into these cells and the localization of endogenous PAR4, GFP-THAP1 and HA-PML (or HA-SP100) was analyzed by triple staining confocal microscopy.

Human Hela cells (ATCC) were grown in Dulbecco's Modified Eagle's Medium supplemented with 10% Fetal Calf Serum and 1% Penicillin-streptomycin (all from Life Technologies, Grand Island, NY, USA), plated on coverslips, and transiently transfected with calcium phosphate method using 2 μg pEGFP.C2-THAP1 and pcDNA.3-HA-PML3 or pSG5-HA-Sp100 (a gift from Dr Dejean, Institut Pasteur, Paris, France) plasmid DNA. pcDNA.3-HA-PML3 was constructed by sub-cloning a BglII-BamHI fragment from pGADT7-HA-PML3 into the BamHI site of pcDNA3 expression vector (Invitrogen, San Diego, CA, USA). To generate pGADT7-HA-PML3, PML3 ORF was amplified by PCR, using pACT2-PML3 (a gift from Dr De Thé, Paris, France) as template, with primers

PML-1 (5'-GCGGGATCCCTAAATTAGAAAGGGGTGGGGGTAGCC-3') (SEQ ID NO: 210) and

PML-2 (5'-GCGGAATTCATGGAGCCTGCACCCGCCCGATC-3') (SEQ ID NO: 211), and cloned into the EcoRI and BamHI sites of pGADT7.

Hela cells transfected with GFP-tagged and HA-tagged expression constructs were allowed to grow for 24 h to 48 h on coverslips. Cells were washed twice with PBS, fixed in methanol for 5 min at –20°C, followed by incubation in cold acetone at –20°C for 30 sec. Permeabilized cells were then blocked with PBS-BSA (PBS with 1% bovine serum albumin) for 10' and then incubated 2 hr at room temperature with the following primary antibodies diluted in PBS-BSA: rabbit polyclonal antibodies against human Par4 (1/50, R-334, Santa Cruz Biotechnology, Santa Cruz, CA, USA) and mouse monoclonal antibody anti-HA tag (mouse IgG1, 1/1000, mAb 16B12 from BabCO, Richmond, CA, USA). Cells were then washed three times 5 min at room temperature in PBS-BSA, and incubated for 1 hr with Cy3 (red fluorescence)-conjugated goat anti-rabbit IgG (1/1000, Amersham Pharmacia Biotech) and Alexa Fluor-633 (blue fluorescence) goat anti-mouse IgG conjugate (1/100, Molecular Probes, Eugene, OR, USA) secondary antibodies, diluted in PBS-BSA. After extensive washing in PBS, samples were air dried and mounted in Mowiol. Images were collected on a Leica confocal laser scanning microscope. The GFP (green), Cy3 (red) and Alexa 633 (blue) fluorescence signals were recorded sequentially for identical image fields to avoid cross-talk between the channels.

In Hela cells transfected with HA-PML, endogenous PAR4 and GFP-THAP1 were recruited to PML-NBs, whereas in cells transfected with HA-SP100, both PAR4 and GFP-THAP1 exhibited diffuse staining without accumulation in PML-NBs. These findings indicate that recruitment of the THAP1/PAR4 complex to PML-NBs depends on PML but not SP100.

EXAMPLE 10

THAP1 is an apoptosis inducing polypeptide

THAP1 is a novel proapoptotic factor

Since PML and PML-NBs have been linked to regulation of cell death and PAR4 is a well established pro-apoptotic factor, we examined whether THAP1 can modulate cell survival. Mouse 3T3 cells, which have previously been used to analyze the pro-apoptotic activity of PAR4 (Diaz-Meco et al , 1996 ; Berra et al., 1997), were transfected with expression vectors for GFP-THAP1, GFP-PAR4 and as a negative control GFP-APS kinase-1 (APSK-1), a nuclear enzyme unrelated to THAP1 and

apoptosis (Girard et al., 1998 ; Besset et al., 2000). We then determined whether ectopic expression of THAP1 enhances the apoptotic response to serum withdrawal. Transfected cells were deprived of serum for up to twenty four hours and cells with apoptotic nuclei, as revealed by DAPI staining and in situ TUNEL assay, were counted.

Cell death assays: Mouse 3T3-TO fibroblasts were seeded on coverslips in 12-well plates at 40 to 50% confluency and transiently transfected with GFP or GFP-fusion protein expression vectors using Lipofectamine Plus reagent (Life Technologies) according to supplier's instructions. After 6h at 37°C, the DNA-lipid mixture was removed and the cells were allowed to recover in complete medium for 24 h. Serum starvation of transiently transfected cells was induced by changing the medium to 0% serum, and the amount of GFP-positive apoptotic cells was assessed 24 h after induction of serum starvation. Cells were fixed in PBS containing 3.7% formaldehyde and permeabilized with 0.1% Triton-X100 as described under immunofluorescence, and apoptosis was scored by *in situ* TUNEL (terminal deoxynucleotidyl transferase-mediated dUTP nick end labeling) and/or DAPI (4,6-Diamidino-2-phenylindole) staining of apoptotic nuclei exhibiting nuclear condensation. The TUNEL reaction was performed for 1 hr at 37°C using the *in situ* cell death detection kit, TMR red (Roche Diagnostics, Meylan, France). DAPI staining with a final concentration of 0.2 :g/ml was performed for 10 min at room temperature. At least 100 cells were scored for each experimental point using a fluorescence microscope.

Basal levels of apoptosis in the presence of serum ranged from 1-3 %. Twenty four hours after serum withdrawal, apoptosis was found in 18% of untransfected 3T3 cells and in 3T3 cells overexpressing GFP-APSK-1.Levels of serum withdrawal induced apoptosis were significantly increased to about 70% and 65% in cells overexpressing GFP-PAR4 and GFP-THAP1, respectively (Figure 6A). These results demonstrate that THAP1, similarly to PAR4, is an apoptosis inducing polypeptide.

TNFα-induced apoptosis assays were performed by incubating transiently transfected cells in complete medium containing 30 ng/ml of mTNFα (R & D, Minneapolis, MN, USA) for 24 h. Apoptosis was scored as described for serum withdrawal-induced apoptosis. The results are shown in Figure 6B. As shown in Figure 6B, THAP1 induced apoptosis.

EXAMPLE 11

The THAP domain is essential for THAP1 pro-apoptotic activity

To determine the role of the amino-terminal THAP domain (amino acids 1 to 89) in the functional activity of THAP1, we generated a THAP1 mutant that is deleted of the THAP domain (THAP1ΔTHAP). THAP1ΔTHAP (amino acids 90-213) was amplified by PCR, using pEGFP.C2-THAP1 as template, with primers *2HMR12* (5'-GCGGAATTCAAAGAAGATCTTCTGGAGCCACAGGAAC-3') (SEQ ID NO: 212) and *2HMR9* (5'-CGCGGATCCTGCTGGTACTTCAACTATTTCAAAGTAGTC-3') (SEQ ID NO: 213), digested with EcoRI and BamHI, and cloned in pGBKT7 and pEGFP-C2 vectors, to generate pGBKT7-THAP1ΔTHAP and pEGFP.C2-THAP1ΔTHAP expression vectors. The role of the THAP domain in PML NBs localization, binding to Par4, or pro-apoptotic activity of THAP1 was then analyzed.

To analyze the subcellular localization of THAP1ΔTHAP, the GFP/ THAP1ΔTHAP expression construct was transfected into human primary endothelial cells from umbilical vein (HUVEC, PromoCell, Heidelberg, Germany). HUVEC were grown in complete ECGM medium (PromoCell, Heidelberg, Germany), plated on coverslips and transiently transfected in RPMI medium using GeneJammer transfection reagent according to manufacturer instructions (Stratagene, La Jolla, CA, USA). Transfected cells were allowed to grow for 48 h on coverslips. Cells were then washed twice with PBS, fixed for 15 min at room temperature in PBS containing 3.7% formaldehyde, and washed again with PBS prior to neutralization with 50mM NH_4Cl in PBS for 5 min at room temperature. Following one more PBS wash, cells were permeabilized 5 min at room temperature in PBS containing 0.1% Triton-X100, and washed again with PBS. Permeabilized cells were then blocked with PBS-BSA (PBS with 1% bovine serum albumin) for 10' and then incubated 2 hr at room temperature with mouse monoclonal antibody anti-PML (mouse IgG1, 1/30, mAb PG-M3 from Dako, Glostrup, Denmark) diluted in PBS-BSA. Cells were then washed three times 5 min at room temperature in PBS-BSA, and incubated for 1 hr with Cy3 (red fluorescence)-conjugated goat anti-mouse IgG (1/1000, Amersham Pharmacia Biotech) secondary antibodies, diluted in PBS-BSA. After extensive washing in PBS, samples were air dried and mounted in Mowiol. Images were collected on a Leica confocal laser scanning microscope. The GFP (green) and Cy3 (red) fluorescence signals were

recorded sequentially for identical image fields to avoid cross-talk between the channels.

This analysis revealed that GFP- THAP1ΔTHAP staining exhibits a complete overlap with the staining pattern obtained with antibodies directed against PML, indicating the THAP domain is not required for THAP1 localization to PML NBs.

To examine the role of the THAP domain in binding to Par4, we performed *in vitro* GST pull down assays. Par4DD, expressed as a GST-tagged fusion protein and immobilized on glutathione sepharose, was incubated with radiolabeled *in vitro* translated THAP1ΔTHAP. *In vitro*-translated THAP1ΔTHAP was generated with the TNT-coupled reticulocyte lysate system (Promega, Madison, WI, USA) using pGBKT7-THAP1ΔTHAP vector as template. 25 µl of ^{35}S-labelled THAP1)ΔTHAP was incubated with immobilized GST-Par4 or GST proteins overnight at 4 °C, in the following binding buffer : 10 mM NaPO4 pH 8.0, 140 mM NaCl, 3 mM MgCl2, 1mM dithiothreitol (DTT), 0.05% NP40, and 0.2 mM phenylmethyl sulphonyl fluoride (PMSF), 1 mM Na Vanadate, 50mM β Glycerophosphate, 25 µg/ml chimotrypsine, 5 µg/ml aprotinin, 10 µg/ml Leupeptin. Beads were then washed 5 times in 1 ml binding buffer. Bound proteins were eluted with 2X Laemmli SDS-PAGE sample buffer, fractionated by 10% SDS–PAGE and visualized by fluorography using Amplify (Amersham Pharmacia Biotech).

This analysis revealed that THAP1ΔTHAP interacts with GST/Par4DD, indicating that the THAP domain is not involved in THAP1/Par4 interaction (Figure 7A).

To examine the role of the THAP domain in THAP1 pro-apoptotic activity, we performed cell death assays in mouse 3T3 cells. Mouse 3T3-TO fibroblasts were seeded on coverslips in 12-well plates at 40 to 50% confluency and transiently transfected with GFP-APSK1, GFP-THAP1 or GFP-THAP1ΔTHAP fusion proteins expression vectors using Lipofectamine Plus reagent (Life Technologies) according to supplier's instructions. After 6h at 37°C, the DNA-lipid mixture was removed and the cells were allowed to recover in complete medium for 24 h. Serum starvation of transiently transfected cells was induced by changing the medium to 0% serum, and the amount of GFP-positive apoptotic cells was assessed 24 h after induction of serum starvation. Cells were fixed in PBS containing 3.7% formaldehyde and permeabilized

with 0.1% Triton-X100 as described under immunofluorescence, and apoptosis was scored by *in situ* TUNEL (terminal deoxynucleotidyl transferase-mediated dUTP nick end labeling) and/or DAPI (4,6-Diamidino-2-phenylindole) staining of apoptotic nuclei exhibiting nuclear condensation. The TUNEL reaction was performed for 1 hr at 37°C using the *in situ* cell death detection kit, TMR red (Roche Diagnostics, Meylan, France). DAPI staining with a final concentration of 0.2 μg/ml was performed for 10 min at room temperature. At least 100 cells were scored for each experimental point using a fluorescence microscope.

Twenty four hours after serum withdrawal, apoptosis was found in 18% of untransfected 3T3 cells and in 3T3 cells overexpressing GFP-APSK-1. Levels of serum withdrawal induced apoptosis were significantly increased to about 70% in cells overexpressing GFP-THAP1. Deletion of the THAP domain abrogated most of this effect since serum-withdrawal-induced apoptosis was reduced to 28 % in cells overexpressing GFP-THAP1ΔTHAP (Figure 7B). These results indicate that the THAP domain, although not required for THAP1 PML-NBs localization and Par4 binding, is essential for THAP1 pro-apoptotic activity.

EXAMPLE 12

The THAP domain defines a novel family of proteins, the THAP family

To discover novel human proteins homologous to THAP1 and/or containing THAP domains, GenBank non-redundant, human EST and draft human genome databases at the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) were searched with both the nucleotide and amino acid sequences of THAP1, using the programs BLASTN, TBLASTN and BLASTP (Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* 215: 403-410). This initial step enabled us to identify 12, distinct human THAP-containing, proteins (hTHAP0 to hTHAP11; Figure 8). In the case of the partial length sequences, assembly of overlapping ESTs together with GENESCAN (Burge, C.and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 268: 78-94) and GENEWISE (Jareborg, N., Birney, E.and Durbin, R. (1999). Comparative analysis of noncoding regions of 77 orthologous mouse and human gene pairs. *Genome Res* 9: 815-824) gene predictions on

the corresponding genomic DNA clones, was used to define the full length human THAP proteins as well as their corresponding cDNAs and genes. CLUSTALW (Higgins, D. G., Thompson, J. D. and Gibson, T. J. (1996). Using CLUSTAL for multiple sequence alignments. *Methods Enzymol* 266: 383-402) was used to carry out the alignment of the 12 human THAP domains with the DNA binding domain of Drosophila P-element transposase (Lee, C. C., Beall, E. L., and Rio, D. C. (1998) *Embo J.* **17**:4166-74), which was colored using the computer program Boxshade (www.ch.embnet.org/software/BOX_form.html) (see Figures 9A and 9B). Equivalent approach to the one described above was used in order to identify the mouse, rat, pig, and various other orthologs of the human THAP proteins (Figure 9C). Altogether, the *in silico* and experimental approaches led to the discovery of 12 distinct human members (hTHAP0 to hTHAP11) of the THAP family of pro-apoptotic factors (Figure 8).

EXAMPLE 13

THAP2 and THAP3 interact with Par-4

To assess whether THAP2 and THAP3 are able to interact with Par-4, yeast two hybrid assays using Par-4 wild type bait (Figure 10B) and *in vitro* GST pull down assays (Figure 10C), were performed as described above (Examples 4 and 5). As shown in Figures 10B and 10C, THAP2 and THAP3 are able to interact with Par-4. A sequence alignment showing the comparison of the THAP domain and the PAR4-binding domain between THAP1, THAP2 and THAP3 is shown in Figure 10A.

EXAMPLE 14

THAP2 and THAP3 are able to induce apoptosis

Serum-induced or TNFα apoptosis analyses were performed as described above (Example 10) in cells transfected with GFP-APSK1, GFP-THAP2 or GFP-THAP3 expression vectors. Apoptosis was quantified by DAPI staining of apoptotic nuclei 24 hours after serum withdrawal or addition of TNFα. The results are shown in Figure 11A (serum withdrawal) and Figure 11B (TNFα). These results indicate that, THAP-2 and THAP3 induce apoptosis.

EXAMPLE 15

Identification of the SLC/CCL21 chemokine-binding domain of human THAP1

To identify the SLC/CCL21 chemokine-binding domain of human THAP1, a series of THAP1 deletion constructs was generated as described in Example 7.

Two-hybrid interaction between THAP1 mutants and chemokine SLC/CCL21 was tested by cotransformation of AH109 with pGADT7-THAP1-C1, -C2, -C3, -N1, -N2 or -N3 and pGBKT7-SLC/CCL21 and selection of transformants by His and Ade double auxotrophy according to manufacturer's instructions (MATCHMAKER two-hybrid system 3, Clontech). pGBKT7-SLC/CCL21 vector was generated by subcloning the BamHI SLC/CCL21 fragment from pGBT9-SLC (see example 1) into the unique BamHI cloning site of vector pGBKT7 (Clontech). Positive two-hybrid interaction with chemokine SLC/CCL21 was observed with mutants THAP1-C1, -C2, -C3, but not with mutants THAP1-N1, -N2 and –N3, suggesting that the SLC/CCL21 chemokine-binding domain of human THAP1 is found between THAP1 residues 143 and 213 (Figure 12).

EXAMPLE 16

In vitro THAP1/chemokine SLC-CCL21 interaction assay

To confirm the interaction observed in yeast two-hybrid system, we performed *in vitro* GST pull down assays. THAP1, expressed as a GST-tagged fusion protein and immobilized on glutathione sepharose, was incubated with radiolabeled *in vitro* translated SLC/CCL21.

To generate the GST-THAP1 expression vector, the full-length coding region of THAP1 (amino acids 1-213) was amplified by PCR from HEVEC cDNA with primers *2HMR8* (5'-CGCGGATCCGTGCAGTCCTGCTCCGCCTACGGC-3') (SEQ ID NO: 214) and *2HMR11* (5'-CCGAATTCTTATGCTGGTACTTCAACTATTTCAAAGTAG-3') (SEQ ID NO: 215), digested with BamHI and EcoRI, and cloned in frame downstream of the Glutathion S-Transferase ORF, between the BamHI and EcoRI sites of the pGEX-2T prokaryotic expression vector (Amersham Pharmacia Biotech, Saclay, France). GST-THAP1 fusion protein encoded by plasmid pGEX-2T-THAP1 and control GST protein encoded by plasmid pGEX-2T, were then expressed in *E. coli* DH5α and purified by affinity chromatography with glutathione sepharose according to supplier's instructions

(Amersham Pharmacia Biotech). The yield of proteins used in GST pull-down assays was determined by SDS-Polyarylamide Gel Electrophoresis (PAGE) and Coomassie blue staining analysis.

In vitro-translated SLC/CCL21 was generated with the TNT-coupled reticulocyte lysate system (Promega, Madison, WI, USA) using as template pGBKT7-SLC/CCL21 vector (see Example 15). 25 µl of ^{35}S-labelled wild-type SLC/CCL21 was incubated with immobilized GST-THAP1 or GST proteins overnight at 4 °C, in the following binding buffer : 10 mM NaPO4 pH 8.0, 140 mM NaCl, 3 mM MgCl2, 1mM dithiothreitol (DTT), 0.05% NP40, and 0.2 mM phenylmethyl sulphonyl fluoride (PMSF), 1 mM Na Vanadate, 50mM β Glycerophosphate, 25 µg/ml chimotrypsine, 5 µg/ml aprotinin, 10 µg/ml Leupeptin. Beads were then washed 5 times in 1 ml binding buffer. Bound proteins were eluted with 2X Laemmli SDS-PAGE sample buffer, fractionated by 10% SDS–PAGE and visualized by fluorography using Amplify (Amersham Pharmacia Biotech). As expected, GST/THAP1 interacted with SLC/CCL21 (Figure 13). In contrast, SLC/CCL21 failed to interact with GST beads.

EXAMPLE 17

Identification of the THAP1-binding domain of human chemokine SLC/CCL21

To determine the THAP1-binding site on human chemokine SLC/CCL21, a SLC/CCL21 deletion mutant (SLC/CCL21ΔCOOH) lacking the SLC-specific basic carboxy-terminal extension (amino acids 102-134 ; GenBank Accession Number NP_002980) was generated. This SLC/CCL21ΔCOOH mutant, which retains the CCR7 chemokine receptor binding domain of SLC/CCL21 (amino acids 24-101), was used both in yeast two-hybrid assays with THAP1 bait and in *in vitro* GST-pull down assays with GST-THAP1.

For two-hybrid assays, yeast cells were cotransformed with BD7-THAP1 and AD7-SLC/CCL21 or AD7-SLC/CCL21ΔCOOH expression vectors. AD7-SLC/CCL21 or AD7-SLC/CCL21ΔCOOH expression vectors were generated by subcloning BamHI fragment (encoding SLC amino acids 24-134) or BamHI-PstI fragment (encoding SLC amino acids 24-102) from pGKT7-SLC/CCL21 (see example 15) into pGADT7 expression vector (Clontech). Transformants were selected on media lacking histidine and adenine. Figure 13 shows that both the SLC/CCL21 wild type and the

SLC/CCL21ΔCOOH deletion mutants could bind to THAP1. Identical results were obtained by cotransformation of AD7-THAP1 with BD7-SLC/CCL21 or BD7-SLC/CCL21ΔCOOH.

GST pull down assays, using *in vitro*-translated SLC/CCL21ΔCOOH, generated with the TNT-coupled reticulocyte lysate system (Promega, Madison, WI, USA) using as template pGBKT7-SLC/CCL21ΔCOOH, were performed as described in Example 16. Figure 13 shows that both the SLC/CCL21 wild type and the SLC/CCL21ΔCOOH deletion mutants could bind to THAP1.

EXAMPLE 18A

Preparation of THAP1/Fc Fusion Proteins

This example describes preparation of a fusion protein comprising THAP1 or the SLC/CCL21 chemokine-binding domain of THAP1 fused to an Fc region polypeptide derived from an antibody. An expression vector encoding the THAP1/Fc fusion protein is constructed as follows.

Briefly, the full length coding region of human THAP1 (SEQ ID NO: 3; amino acids -1 to 213) or the SLC/CCL21 chemokine-binding domain of human THAP1 (SEQ ID NO: 3; amino acids -143 to 213) is amplified by PCR. The oligonucleotides employed as 5' primers in the PCR contain an additional sequence that adds a Not I restriction site upstream. The 3' primer includes an additional sequence that encodes the first two amino acids of an Fc polypeptide, and a sequence that adds a Bgl II restriction site downstream of the THAP1 and Fc sequences.

A recombinant vector containing the human THAP1 cDNA is employed as the template in the PCR, which is conducted according to conventional procedures. The amplified DNA is then digested with Not I and Bgl II, and the desired fragments are purified by electrophoresis on an agarose gel.

A DNA fragment encoding the Fc region of a human IgG1 antibody is isolated by digesting a vector containing cloned Fc-encoding DNA with Bgl II and Not I. Bgl II cleaves at a unique Bgl II site introduced near the 5' end of the Fc-encoding sequence, such that the Bgl II site encompasses the codons for amino acids three and four of the Fc polypeptide. Not I cleaves downstream of the Fc-encoding sequence. The nucleotide sequence of cDNA encoding the Fc polypeptide, along with the encoded

amino acid sequence, can be found in International Publication No: WO93/10151, incorporated herein by reference in its entirety.

In a three-way ligation, the above-described THAP1 (or SLC/CCL21 chemokine-binding domain of THAP1) -encoding DNA and Fc-encoding DNA are inserted into an expression vector that has been digested with Not I and treated with a phosphatase to minimize recircularization of any vector DNA without an insert. An example of a vector which can be used is pDC406 (described in McMahan et al., EMBO J. 10:2821, 1991), which is a mammalian expression vector that is also capable of replication in E. coli.

E. coli cells are then transfected with the ligation mixture, and the desired recombinant vectors are isolated. The vectors encode amino acids-1 to 213 of the THAP1 sequence (SEQ ID NO: 3) or amino acids-143 to 213 of the THAP1 sequence of (SEQ ID NO: 3), fused to the N-terminus of the Fc polypeptide. The encoded Fc polypeptide extends from the N-terminal hinge region to the native C-terminus, i.e., is an essentially full-length antibody Fc region.

CV-1/EBNA-1 cells are then transfected with the desired recombinant isolated from *E. coli*. CV-1/EBNA-1 cells (ATCC CRL 10478) can be transfected with the recombinant vectors by conventional procedures. The CVI-EBNA-1 cell line was derived from the African Green Monkey kidney cell line CV-1 (ATCC CCL 70), as described by McMahan et al. (1991). *EMBO J.* **10**:2821. The transfected cells are cultured to allow transient expression of the THAP1/Fc or SLC/CCL21 chemokine-binding domain of THAP1/Fc fusion proteins, which are secreted into the culture medium. The secreted proteins contain the mature form of THAP1 or the SLC/CCL21 chemokine-binding domain of THAP1, fused to the Fc polypeptide. The THAP1/Fc and SLC/CCL21 chemokine-binding domain of THAP1/Fc fusion proteins are believed to form dimers, wherein two such fusion proteins are joined by disulfide bonds that form between the Fc moieties thereof. The THAP1/Fc and SLC/CCL21 chemokine-binding domain of THAP1/Fc fusion proteins can be recovered from the culture medium by affinity chromatography on a Protein A-bearing chromatography column.

EXAMPLE 18B

Preparation of THAP1/IgG1-Fc Fusion Proteins

This example describes preparation of a fusion protein comprising THAP1 or the SLC/CCL21 chemokine-binding domain of THAP1 fused to an Fc region polypeptide derived from an antibody. An expression vector encoding the THAP1/IgG1-Fc fusion protein was derived from a pCDM8 expression vector encoding L-selectin-IgG1 fusion proteins (recombinant chimeric molecules containing extracellular regions of L-selectin coupled to the hinge, CH2, and CH3 regions of human IgG1) as described in Aruffo, A., et al., Cell, 67:35, 1991, and Walz, G., et al., Science, 250:1132, 1990, the disclosures of which are incorporated herein by reference in their entireties. The nucleotide sequence of cDNA encoding the IgG1-Fc polypeptide, along with the encoded amino acid sequence is described in International Publication No. WO93/10151, the disclosure of which is incorporated herein by reference in its entirety.

Briefly, the full length coding region of human THAP1 (SEQ ID NO: 3; amino acids -2 to 213) or the SLC/CCL21 chemokine-binding domain of human THAP1 (CBD/THAP1, SEQ ID NO: 3; amino acids -140 to 213) were amplified by PCR with primers THAP1-XhoI-5' (5'-CCGCTCGAGGTGCAGTCCTGCT-3') (SEQ ID NO: 264) and THAP1-BamHI-3' (5'-CGGGATCCGCTGGTACTTCAACTATTTCA-3') (SEQ ID NO: 265), or primers CBD/THAP1-XhoI-5' (5'-CCGCTCGAGGATACAATGCACC-3') (SEQ ID NO: 266) and CBD/THAP1-BamH1-3' (5'-GCGGGATCCGCTGGTACTTCAACTATTTCAAAG-3') (SEQ ID NO: 267), respectively. A recombinant vector containing the human THAP1 cDNA (see example 7) was employed as the template in the PCR, which was conducted according to conventional procedures. The amplified DNAs were then digested with Xho I and BamH I and the desired fragments were purified by electrophoresis on an agarose gel. The resulting Xho I-BamH I fragments were then used to replace the Xho I-BamH I fragment encoding L-selectin in the plasmid pCDM8-L-selectin-IgG1 (Aruffo, A., et al., Cell, 67:35, 1991; Walz, G., et al., Science, 250:1132, 1990). The recombinant vectors thus obtained, pCDM8-THAP1-IgG1 and pCDM8-CBD/THAP1-IgG1, encode amino acids-2 to 213 of the THAP1 sequence (SEQ ID NO: 3) or amino acids-140 to 213 of the THAP1 sequence of (SEQ ID NO: 3), fused to the N-terminus

of the IgG1-Fc polypeptide. Because the encoded IgG1-Fc region of the fusion polypeptides extend from the N-terminal hinge region to the native C-terminus, the IgG1-Fc region is essentially a full-length antibody Fc region.

In addition to fusion the IgG1-Fc region to THAP1 and CBD/THAP1, the signal peptide of immunoglobulin kappa light chain was fused to the N-terminus of each of these proteins. A nucleic acid encoding the signal peptide was obtained by using PCR to amplify a SalI-XhoI signal peptide cassette in two steps. In the first step, the oligonucleotide psignal5' (5'-CCGCTCGAG CCACCATGGAGACAGACACACTCCTGCTATGGGTACTGCTGCTCTGGGTTCC AGGTTCCACTGGTGACCTCGAGATT-3') (SEQ ID NO: 268), which encodes the 21 amino acids of the immunoglobulin kappa chain signal peptide from plasmid vector pSecTag2 (Invitrogen) was synthesized and then used as a template for PCR with primers psignal-SalI 5' (5'-TAGGGTCGACGCCACCATGGAGACAG-3') (SEQ ID NO: 269) and psignalXhoI 3' (5'-CCGCTCGAGGTCACCAGTGGA-3') (SEQ ID NO: 270). The product of the PCR reaction was digested with Sal I and Xho I and ligated into the Xho I site of plasmids pCDM8-THAP1-IgG1 and pCDM8-CBD/THAP1-IgG1 to obtain expression vectors pCDM8-SS-THAP1-IgG1 and pCDM8-SS-CBD/THAP1-IgG1. These plasmids were then transfected in COS cells or CV-1/EBNA-1 cells (ATCC CRL 10478), as previously described (Seed, B., et al., Proc. Natl. Acad. Sci., U.S.A., 84:3365, 1987; Aruffo, A., Current Protocols In Molecular Biology, eds. Ausubel, F. M., et al, 16:13.1, Greene Publishing Associates and Wiley-Interscience, New York, N.Y., 1992, the disclosures of which are incorporated herein by reference in their entireties). The CVI-EBNA-1 cell line was derived from the African Green Monkey kidney cell line CV-1 (ATCC CCL 70), as described by McMahan et al. (1991). *EMBO J.* **10**:2821, the disclosure of which is incorporated by reference herein in its entirety. The transfected cells were cultured to allow transient expression of the THAP1/Fc or SLC/CCL21 chemokine-binding domain of THAP1/Fc fusion proteins, which were secreted into the culture medium. The proteins that were secreted contain the mature form of THAP1 or the SLC/CCL21 chemokine-binding domain of THAP1, fused to the Fc polypeptide. Although not bound by therory, the THAP1/Fc and SLC/CCL21 chemokine-binding domain of THAP1/Fc fusion proteins are believed to form dimers, wherein two such fusion proteins are joined by disulfide bonds that form

between the Fc moieties thereof. The THAP1/Fc and SLC/CCL21 chemokine-binding domain of THAP1/Fc fusion proteins were recovered from the culture medium by affinity chromatography on a Protein A-bearing chromatography column.

EXAMPLE 19

The THAP domain defines a family of nuclear factors

To determine the subcellular localization of the different human THAP proteins, a series of GFP-THAP expression constructs were transfected into primary human endothelial cells. In agreement with the possible functions of THAP proteins as DNA-binding factors, we found that all the human THAP proteins analyzed (THAP0, 1, 2, 3, 6, 7, 8, 10, 11) localize preferentially to the cell nucleus (Figure 14). In addition to their diffuse nuclear localization, some of the THAP proteins also exhibited association with distinct subnuclear structures: the nucleolus for THAP2 and THAP3, and punctuate nuclear bodies for THAP7, THAP8 and THAP11. Indirect immunofluorescence microscopy with anti-PML antibodies revealed that the THAP8 and THAP11 nuclear bodies colocalize with PML-NBs. Although the THAP7 nuclear bodies often appeared in close association with the PML-NBs, they never colocalized.

Analysis of the subcellular localization of the GFP-THAP fusion proteins was performed as described above (Example 3). The GFP-THAP constructs were generated as follows: the human THAP0 coding region was amplified by PCR from Hevec cDNA with primers *THAP0-1* (5'-GCCGAATTCATGCCGAACTTCTGCGCTGCCCCC-3') (SEQ ID NO: 216) and *THAP0-2* (5'-CGCGGATCCTTAGGTTATTTTCCACAGTTTCGGAATTATC-3') (SEQ ID NO: 217), digested with EcoRI and BamHI, and cloned in the same sites of the pEGFP-C2 vector, to generate pEGFPC2-THAP0 ; the coding region of human THAP2, 3, 7, 6 and 8 were amplified by PCR respectively from Image clone No: 3606376 with primers *THAP2-1* (5'-GCGCTGCAGCAAGCTAAATTTAAATGAAGGTACTCTTGG-3') (SEQ ID NO: 218) and *THAP2-2* (5'-GCGAGATCTGGGAAATGCCGACCAATTGCGCTGCG-3') (SEQ ID NO: 219) digested with BglII and PstI, from Image clone No: 4813302 and No: 3633743 with primers *THAP3-1* (5'-AGAGGATCCTTAGCTCTGCTGCTCTGGCCCAAGTC-3') (SEQ ID NO: 220) *THAP3-2* (5'-

AGAGAATTCATGCCGAAGTCGTGCGCGGCCCG-3') (SEQ ID NO: 221) and primers *THAP7-1* (5'-GCGGAATTCATGCCGCGTCACTGCTCCGCCGC-3') (SEQ ID NO: 222) *THAP7-2* (5'-GCGGGATCCTCAGGCCATGCTGCTGCTCAGCTGC-3') (SEQ ID NO: 223), digested with EcoRI and BamHI, from Image clone No: 757753 with primers *THAP6-1* (5'-GCGAGATCTCGATGGTGAAATGCTGCTCCGCCATTGGA-3') (SEQ ID NO: 224) and *THAP6-2* (5'-GCGGGATCCTCATGAAATATAGTCCTGTTCTATGCTCTC-3') (SEQ ID NO: 225) digested with BglII and BamHI, and from Image clone No: 4819178 with primers *THAP8-1* (5'-GCGAGATCTCGATGCCCAAGTACTGCAGGGCGCCG-3') (SEQ ID NO: 226) and *THAP8-2* (5'-GCGGAATTCTTATGCACTGGGGATCCGAGTGTCCAGG-3') (SEQ ID NO: 227), digested with BglII and EcoRI and cloned in frame downstream of the Enhanced Green Fluorescent Protein (EGFP) ORF in pEGFPC2 vector (Clontech) digested with the same enzymes to generate pEGFPC2-THAP2, -THAP3, -THAP7, -THAP6 and -THAP8 ; the human THAP10 and THAP11 coding region were amplified by PCR from Hela cDNA respectively with primers *THAP10-1* (5'-GCGGAATTCATGCCGGCCCGTTGTGTGGCCGC-3') (SEQ ID NO: 228) *THAP10-2* (5'-GCGGGATCCTTAACATGTTTCTTCTTTCACCTGTACAGC-3') (SEQ ID NO: 229) digested with EcoRI and BamHI, and with primers *THAP11-1* (5'-GCGAGATCTCGATGCCTGGCTTTACGTGCTGCGTGC-3') (SEQ ID NO: 230) and *THAP11-2* (5'-GCGGAATTCTCACATTCCGTGCTTCTTGCGGATGAC-3') (SEQ ID NO: 231), digested with BglII and EcoRI, cloned in the same sites of the pEGFP-C2 vector, to generate pEGFPC2-THAP10 and –THAP11.

EXAMPLE 20

The THAP domain shares structural similarities with the DNA-binding domain of nuclear hormone receptors

In an effort to model the three-dimensional structure of the THAP domain, we searched the PDB crystallographic database. As sequence homology detection is more sensitive and selective when aided by secondary structure information, structural homologs of the THAP domain of human THAP1 were searched using the SeqFold threading program (Olszewski et al. (1999) Theor. Chem. Acc. 101, 57-61) which

combines sequence and secondary structure alignment. The crystallographic structure of the thyroid hormone receptor β DBD (PDB code: 2NLL) gave the best score of the search and we used the resulting structural alignment, displayed in Figure 15A, to derive a homology-based model of the THAP domain from human THAP1 (Figure 15B). Note that the distribution of Cys residues in the THAP domain does not fully match that of the thyroid hormone receptor β DBD (Figure 15A) and hence cannot allow the formation of the two characteristic 'C4-type' Zn-fingers (red color-coding in Figure 15A). However, a network of stacking interactions between aromatic/hydrophobic residues or aliphatic parts of lysine side-chains ensures the stability of the structure of the THAP domain (cyan color-coding in Figures 15A and 15B). Interestingly the same threading method applied independently to the Drosophila P-element transposase DBD identified the crystallographic structure of the glucocorticoid receptor DBD (PDB code: 1GLU) as giving the best score. In the same way, we used the resulting structural alignment, displayed in Figure 15D, to build a model of the transposase DBD (Figure 15C). Note the presence of an hydrophobic core equivalent to that of the THAP domain (cyan color-coding in Figures 15C and 15D). All the DNA-binding domains of the nuclear receptors fold into a typical pattern which is mainly based on two interacting α-helices, the first one inserting into the target DNA major groove. Our threading and modeling results indicate that the THAP domain and the *D. melanogaster* P-element transposase DBD likely share a common topology which is similar to that of the DBD of nuclear receptors.

Molecular modeling was performed using the InsightII, SeqFold, Homology and Discover modules from the Accelrys (San Diego, CA) molecular modeling software (version 98), run on a Silicon Graphics O2 workstation. Optimal secondary structure prediction of the query protein domains was ensured by the DSC method within SeqFold. The threading–derived secondary structure alignments was used as input for homology-modeling, which was performed according to a previously described protocol (Manival et al. (2001) Nucleic Acids Res 29 :2223-2233). The validity of the models was checked both by Ramachandran analysis and folding consistency verification as previously reported (Manival et al. (2001) Nucleic Acids Res 29 :2223-2233).

EXAMPLE 21

Homodimerization domain of human THAP1

To identify the sequences mediating homodimerization of THAP1, a series of THAP1 deletion constructs was generated as described in Example 7.

Two-hybrid interaction between THAP1 mutants and THAP1 wild type was tested by cotransformation of AH109 with pGADT7-THAP1-C1, -C2, -C3, -N1, -N2 or -N3 and pGBKT7-THAP1 wild-type and selection of transformants by His and Ade double auxotrophy according to manufacturer's instructions (MATCHMAKER two-hybrid system 3, Clontech). Positive two-hybrid interaction with THAP1 wild type was observed with mutants THAP1-C1, -C2, -C3, -and -N3 but not with mutants THAP1-N1 and -N2, suggesting the THAP1 homodimerization domain is found between THAP1 residues 143 and 192 (Figure 16A).

To confirm the results obtained in yeast, THAP1 mutants were also tested in *in vitro* GST pull down assays. Wild type THAP1 expressed as a GST-tagged fusion protein and immobilized on glutathione sepharose (as described in example 16), was incubated with radiolabeled *in vitro* translated THAP1 mutants. *In vitro*-translated THAP1 mutants were generated with the TNT-coupled reticulocyte lysate system (Promega, Madison, WI, USA) using pGADT7-THAP1-C1, -C2, -C3, -N1, -N2 or -N3 vector as template. 25 µl of each ^{35}S-labelled THAP1 mutant was incubated with immobilized GST or GST-THAP1 wild-type protein overnight at 4 °C, in the following binding buffer : 10 mM NaPO4 pH 8.0, 140 mM NaCl, 3 mM MgCl2, 1mM dithiothreitol (DTT), 0.05% NP40, and 0.2 mM phenylmethyl sulphonyl fluoride (PMSF), 1 mM Na Vanadate, 50mM β Glycerophosphate, 25 µg/ml chimotrypsine, 5 µg/ml aprotinin, 10 µg/ml Leupeptin. Beads were then washed 5 times in 1 ml binding buffer. Bound proteins were eluted with 2X Laemmli SDS-PAGE sample buffer, fractionated by 10% SDS–PAGE and visualized by fluorography using Amplify (Amersham Pharmacia Biotech). As expected, THAP1-C1, -C2, -C3, -and -N3 interacted with GST/THAP1 (Figure 16B). In contrast, THAP1-N1 and -N2 failed to interact with GST/THAP1 beads. Therefore, essentially identical results were obtained with the two THAP1/THAP1 interactions assays: the THAP1 homodimerization domain of THAP1 is found between residues 143 and 192 of human THAP1.

EXAMPLE 22

Alternatively spliced isoform of human THAP1

The two distinct THAP1 cDNAs, THAP1a and THAP1b have been discovered (Figure 17A). These splice variants, were amplified by PCR from HEVEC cDNA with primers *2HMR10* (5'-CCGAATTCAGGATGGTGCAGTCCTGCTCCGCCT-3') (SEQ ID NO: 232) and *2HMR9* (5'-CGCGGATCCTGCTGGTACTTCAACTATTTCAAAGTAGTC-3') (SEQ ID NO: 233), digested with EcoRI and BamHI, and cloned in frame upstream of the Enhanced Green Fluorescent Protein (EGFP) ORF in pEGFP.N3 vector (Clontech) to generate pEGFP.N3-THAP1a and pEGFP-THAP1b. DNA sequencing revealed that THAP1b cDNA isoform lacks exon 2 (nucleotides 273-468) of the human THAP1 gene (Figure 17B). This alternatively spliced isoform of human THAP1 (~ 2 kb mRNA) was also observed in many other tissues by Northern blot analysis (see Figure 2). The THAP1a/GFP and THAP1b/GFP expression constructs were then transfected into COS 7 cells (ATCC) and expression of the fusion proteins was analyzed by western blotting with anti-GFP antibodies. The results are shown in Figure 17C which demonstrates that the second isoform of human THAP1 (THAP1b) encodes a truncated THAP1 protein (THAP1 C3) lacking a substantial portion of the amino terminus (amino acids 1-142 of SEQ ID NO: 3).

EXAMPLE 23

High throughput screening assay for modulators of THAP family Polypeptide pro-apoptotic activity

A high throughput screening assay for molecules that abrogate or stimulate THAP-family polypeptide proapoptotic activity was developed, based on serum-withdrawal induced apoptosis in a 3T3 cell line with tetracycline-regulated expression of a THAP family polypeptide.

In a preferred example, the THAP1 cDNA with an in-frame myc tag sequence, was amplified by PCR using pGBKT7-THAP1 as a template with primers *myc.BD7* (5'-GCGCTCTAGAGCCATCATGGAGGAGCAGAAGCTGATC-3') (SEQ ID NO: 234) and *2HMR15* (5'-GCGCTCTAGATTATGCTGGTACTTCAACTATTTCAAAGTAG-3') (SEQ ID NO: 235), and cloned downstream of a tetracycline regulated promoter in plasmid vector pTRE (Clontech, Palo Alto, CA), using *Xba I* restriction site, to generate

plasmid pTRE-mycTHAP1. To establish 3T3-TO-mycTHAP1 stable cell lines, mouse 3T3-TO fibroblasts (Clontech) were seeded at 40 to 50% confluency and co-transfected with the pREP4 plasmid (Invitrogen), which contains a hygromycin B resistance gene, and the mycTHAP1 expression vector (pTRE-mycTHAP1) at 1:10 ratio, using Lipofectamine Plus reagent (Life Technologies) according to supplier's instructions. Transfected cells were selected in medium containing hygromycin B (250 U/ml; Calbiochem) and tetracycline (2 ug/ml; Sigma). Several resistant colonies were picked and analyzed for the expression of mycTHAP1 by indirect immunofluorescence using anti-myc epitope monoclonal antibody (mouse IgG1, 1/200, Clontech). A stable 3T3-TO cell line expressing mycTHAP1 (3T3-TO-mycTHAP1) was selected and grown in Dulbecco's Modified Eagle's Medium supplemented with 10% Fetal Calf Serum, 1% Penicillin-streptomycin (all from Life Technologies, Grand Island, NY, USA) and tetracycline (2 ug/ml; Sigma). Induction of THAP1 expression into this 3T3-TO-mycTHAP1 cell line was obtained 48 h after removal of tetracycline in the complete medium.

A drug screening assay using the 3T3-TO-mycTHAP1 cell line can be carried out as follows. 3T3-TO-mycTHAP1 cells are plated in 96- or 384-wells microplates and THAP1 expression is induced by removal of tetracycline in the complete medium. 48 h later, the apoptotic response to serum withdrawal is assayed in the presence of a test compound, allowing the identification of test compounds that either enhance or inhibit the ability of THAP1 polypeptide to induce apoptosis. Serum starvation of 3T3-TO-mycTHAP1 cells is induced by changing the medium to 0% serum, and the amount of cells with apoptotic nuclei is assessed 24 h after induction of serum starvation by TUNEL labeling in 96- or 384-wells microplates. Cells are fixed in PBS containing 3.7% formaldehyde and permeabilized with 0.1% Triton-X100, and apoptosis is scored by *in situ* TUNEL (terminal deoxynucleotidyl transferase-mediated dUTP nick end labeling) staining of apoptotic nuclei for 1 hr at 37°C using the *in situ* cell death detection kit, TMR red (Roche Diagnostics, Meylan, France). The intensity of TMR red fluorescence in each well is then quantified to identify test compounds that modify the fluorescence signal and thus either enhance or inhibit THAP1 pro-apoptotic activity.

EXAMPLE 24

High throughput two-hybrid screening assay for drugs that modulate THAP-family polypeptide/THAP-family target protein interaction

To identify drugs that modulate THAP1/Par4 or THAP1/SLC interactions, a two-hybrid based high throughput screening assay can be used.

As described in Example 17, AH109 yeast cells (Clontech) cotransformed with plasmids pGBKT7-THAP1 and pGADT7-Par4 or pGADT7-SLC can be grown in 384-well plates in selective media lacking histidine and adenine, according to manufacturer's instructions (MATCHMAKER two-hybrid system 3, Clontech).

Growth of the transformants on media lacking histidine and adenine is absolutely dependent on the THAP1/Par4 or THAP1/SLC two-hybrid interaction and drugs that disrupt THAP1/Par4 or THAP1/SLC binding will therefore inhibit yeast cell growth.

Small molecules (5 mg ml^{-1} in DMSO; Chembridge) are added by using plastic 384-pin arrays (Genetix). The plates are incubated for 4 to 5 days at 30 °C, and small molecules which inhibit the growth of yeast cells by disrupting THAP1/Par4 or THAP1/SLC two-hybrid interaction are selected for further analysis.

EXAMPLE 25

High throughput in vitro assay to identify inhibitors of THAP-family polypeptide/THAP-family protein target interaction

To identify small molecule modulators of THAP function, a high-throughput screen based on fluorescence polarization (FP) is used to monitor the displacement of a fluorescently labelled THAP1 protein from a recombinant glutathione-S-transferase (GST)-THAP binding domain of Par4 (Par4DD) fusion protein or a recombinant GST-SLC/CCL21 fusion protein.

Assays are carried out essentially as in Degterev et al, Nature Cell Biol. 3: 173-182 (2001) and Dandliker et al, Methods Enzymol. 74: 3-28 (1981). The assay can be calibrated by titrating a THAP1 peptide labelled with Oregon Green with increasing amounts of GST-Par4DD or GST-SLC/CCL21 proteins. Binding of the peptide is accompanied by an increase in polarization (mP, millipolarization).

THAP 1 and PAR4 polypeptides and GST-fusions can be produced as previously described. The THAP1 peptide was expressed and purified using a QIAexpressionist kit (Qiagen) according to the manufacturer's instructions. Briefly, the entire THAP1 coding sequence was amplified by PCR using pGBKT7-THAP1 as a template with primers *2HMR8* (5'-CGCGGATCCGTGCAGTCCTGCTCCGCCTACGGC-3') (SEQ ID NO: 236) and *2HMR9* (5'-CGCGGATCCTGCTGGTACTTCAACTATTTCAAAGTAGTC-3') (SEQ ID NO: 237), and cloned into the *Bam*HI site of pQE30 vector (Qiagen). The resulting pQE30-HisTHAP1 plasmid was transformed in *E.coli* strain M15 (Qiagen). 6xHis-tagged-THAP1 protein was purified from inclusion bodies on a Ni-Agarose column (Qiagen) under denaturing conditions, and the eluate was used for in vitro interaction assays. To produce GST-Par4DD fusion protein, Par4DD (amino acids 250-342) was amplified by PCR with primers *Par4.10* (5'-GCCGGATCCGGGTTCCCTAGATATAACAGGGATGCAA-3') (SEQ ID NO: 238) and *Par4.5* (5'-GCGGGATCCCTCTACCTGGTCAGCTGACCCACAAC-3') (SEQ ID NO: 239), and cloned in frame downstream of the Glutathione S-Transferase (GST) ORF, into the BamHI site of the pGEX-2T prokaryotic expression vector (Amersham Pharmacia Biotech, Saclay, France). Similarly, to produce GST-SLC/CCL21 fusion protein, the mature form of human SLC/CCL21 (amino acids 24-134) was amplified by PCR with primers hSLCbam.5' (5'-GCGGGATCCAGTGATGGAGGGGCTCAGGACTGTTG-3') (SEQ ID NO: 240) and hSLCbam.3' (5'-GCGGGATCCCTATGGCCCTTTAGGGGTCTGTGACC-3') (SEQ ID NO: 241), digested with BamHI and inserted into the BamHI cloning site of the pGEX-2T vector. GST-Par4DD (amino acids 250-342) and GST-SLC/CCL21 (amino acids 24-134) fusion proteins were expressed in *E.Coli* DH5α (supE44, DELTAlacU169 (80lacZdeltaM15), hsdR17, recA1, endA1, gyrA96, thi1, relA 1) and purified by affinity chromatography with glutathione sepharose according to supplier's instructions (Amersham Pharmacia Biotech).

For screening small molecules, THAP1 peptide is labelled with succinimidyl Oregon Green (Molecular Probes, Oregon) and purified by HPLC. 33 nM labeled THAP1 peptide, 2μM GST-Par4DD or GST-SLC/CCL21 protein, 0.1% bovine gamma-globulin (Sigma) and 1 mM dithiothreitol mixed with PBS, pH 7.2 (Gibco), are

added to 384-well black plates (Lab Systems) with Multidrop (Lab Systems). Small molecules (5 mg ml^{-1} in DMSO; Chembridge) are transferred by using plastic 384-pin arrays (Genetix). The plates are incubated for 1-2 hours at 25 °C, and FP values are determined with an Analyst plate reader (LJL Biosystems).

EXAMPLE 26

High throughput chip assay to identify inhibitors of THAP-family polypeptide/THAP-family protein target interaction

A chip based binding assay Degterev et al, (2001) Nature Cell Biol. 3: 173-182 using unlabelled THAP and THAP-family target protein may be used to identify molecules capable of interfering with THAP-family and THAP-family target interactions, providing high sensitivity and avoiding potential interference from label moieties. In this example, the THAP1 binding domain of Par4 protein (Par4DD) or SLC/CCL21 is covalently attached to a surface-enhanced laser desorption/ionization (SELDI) chip, and binding of unlabelled THAP1 protein to immobilized protein in the presence of a test compound is monitored by mass spectrometry.

Recombinant THAP1 protein, GST-Par4DD and GST-SLC/CCL21 fusion proteins are prepared as described in Example 25. Purified recombinant GST-Par4DD or GST-SLC/CCL21 protein is coupled through its primary amine to SELDI chip surfaces derivatized with cabonyldiimidazole (Ciphergen). THAP1 protein is incubated in a total volume of 1 µl for 12 hours at 4 °C in a humidified chamber to allow binding to each spot of the SELDI chip, then washed with alternating high-pH and low-pH buffers (0.1M sodium acetate containing 0.5M NaCl, followed by 0.01 M HEPES, pH 7.3). The samples are embedded in an alpha-cyano-4-hydroxycinnamic acid matrix and analysed for mass by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry. Averages of 100 laser shots at a constant setting are collected over 20 spots in each sample.

EXAMPLE 27

High throughput cell assay to identify inhibitors of THAP-family polypeptide/THAP-family protein target interaction

A fluorescence resonance energy transfer (FRET) assay is carried out between THAP-1 and PAR4 or SLC/CCL21 proteins fused with fluorescent proteins. Assays can be carried out as in Majhan et al, Nature Biotechnology 16: 547-552 (1998) and Degterev et al, Nature Cell Biol. 3: 173-182 (2001).

THAP-1 protein is fused to cyan fluorescent protein (CFP) and PAR4 or SLC/CCL21 protein is fused to yellow fluorescent protein (YFP). Vectors containing THAP-family and THAP-family target proteins can be constructed essentially as in Majhan et al (1998). A THAP-1-CFP expression vector is generated by subcloning a THAP-1 cDNA into the pECFP-N1 vector (Clontech). PAR4-YFP or SLC/CCL21-CYP expression vectors are generated by subcloning a PAR4 or a SLC/CCL21 cDNA into the pEYFP-N1 vector (Clontech).

Vectors are cotransfected to HEK-293 cells and cells are treated with test compounds. HEK-293 cells are transfected with THAP-1-CFP and PAR4-YFP or SLC/CCL21-YFP expression vectors using Lipofect AMINE Plus (Gibco) or TransLT-1 (PanVera). 24 hours later cells are treated with test compounds and incubated for various time periods, preferably up to 48 hours. Cells are harvested in PBS, optionally supplemented with test compound, and fluorescence is determined with a C-60 fluorimeter (PTI) or a Wallac plate reader. Fluorescence in the samples separately expressing THAP-1-CFP and PAR4-YFP or SLC/CCL21-YFP is added together and used to estimate the FRET value in the absence of THAP-1/PAR4 or THAP1/SLC/CCL21 binding.

The extent of FRET between CFP and YFP is determined as the ratio between the fluorescence at 527 nm and that at 475 nm after excitation at 433 nm. The cotransfection of THAP-1 protein and PAR4 or SLC/CCL21 protein results in an increase of FRET ratio over a reference FRET ratio of 1.0 (determined using samples expressing the proteins separately). A change in the FRET ratio upon treatmemt with a test compound (over that observed after cotransfection in the absence of a test

compound) indicates a compound capable of modulating the interaction of the THAP-1 protein and the PAR4 or the SLC/CCL21 protein.

EXAMPLE 28

In vitro assay to identify THAP-family polypeptide DNA targets

DNA binding specificity of THAP1 was determined using a random oligonucleotide selection method allowing unbiased analysis of binding sites selected by the THAP domain of the THAP1 protein from a random pool of possible sites. The method was carried out essentially as described in Bouvet (2001) *Methods Mol Biol.* **148**:603-10. Also, see Pollack and Treisman (1990) *Nuc. Acid Res.* **18**:6197-6204; Blackwell and Weintraub, (1990) *Science* **250**: 1104-1110; Ko and Engel, (1993) *Mol. Cell. Biol.* **13**:4011-4022; Merika and Orkin, (1993) *Mol. Cell. Biol.* **13**: 3999-4010; and Krueger and Morimoto, (1994) *Mol. Cell. Biol.* **14**:7592-7603), the disclosures of which are incorporated herein by reference in their entireties.

Recombinant THAP domain expression and purification

A cDNA fragment encoding the THAP domain of human THAP-1 (amino acids 1-90, SEQ ID NO: 3) was cloned by PCR using as a template pGADT7-THAP-1 (see Example 4) with the following primers 5'-GCGCATATGGTGCAGTCCTGCTCCGCCTACGGC-3' (SEQ ID NO: 242) and 5'-GCGCTCGAGTTTCTTGTCATGTGGCTCAGTACAAAG-3' (SEQ ID NO: 243). The PCR product was cloned as a NdeI-XhoI fragment into pET-21c prokaryotic expression vector (Novagen) in frame with a sequence encoding a carboxy terminal His tag, to generate pET-21c-THAP.

For the expression of THAP-His6, pET-21c-THAP was transformed into Escherichia coli strain BL-21 pLysS. Bacteria were grown at 37°C to an optical density at 600nm of 0.6 and expression of the protein was induced by adding isopropyl-β-D-thiogalactoside (Sigma) at a final concentration of 1mM and incubation was continued for 4 hours.

The cells were collected by centrifugation and resuspended in ice cold of buffer A (50 mM sodium-phosphate pH 7.5, 300 mM NaCl, 0.1% β-mercaptoethanol, 10 mM Imidazole). Cells were lysed by sonication and the lysate was cleared by centrifugation

at 12000g for 45 min. The supernatant was loaded onto a Ni-NTA agarose column (Quiagen) equilibrated in buffer A. After washing with buffer A and Buffer A with 40 mM Imidazole, the protein was eluted with buffer B (same as A with 0.05%β-mercaptoethanol and 250 mM Imidazole).

Fractions containing THAP-His6 were pooled and applied to a Superdex 75 gel filtration column equilibrated in Buffer C (Tris-HCl 50mM pH 7.5, 150 mM NaCl, 1 mM DTT). Fractions containing the THAP-His6 protein were pooled, concentrated with YM-3 Amicon filter devices and stored at 4°C or frozen at –80°C in buffer C containing 20% glycerol. The purity of the sample was assessed by SDS-Polyarylamide Gel Electrophoresis (PAGE) and Coomassie blue staining analysis. The structural integrity of the protein preparation was checked by ESI mass spectrometry and Peptide mass mapping using a MALDI-TOF Mass spectrometer.The protein concentration was determined with Bradford Protein Assay.

Random Oligonucleotide Selection

According to the SELEX protocol described in Bouvet (2001) *Methods Mol Biol.* **148**:603-10, a 62 bp oligonucleotide having sequences as follows was synthesized: 5'-TGGGCACTATTTATATCAAC-N25-AATGTCGTTGGTGGCCC-3' (SEQ ID NO: 244) where N is any nucleotide, and primers complementary to each end. Primer P is: 5'-ACCGCAAGCTTGGGCACTATTTATATCAAC-3' (SEQ ID NO: 245), and primer R is 5'-GGTCTAGAGGGCCACCAACGCATT-3' (SEQ ID NO: 246). The 62-mer oligonucleotide is made double stranded by PCR using the P and R primers generating a 80 bp random pool.

About 250 ng of THAP-His6 was incubated with Ni-NTA magnetic beads in NT2 buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.05% NP-40) for 30 min at 4°C on a roller. The beads were washed 2 times with 500 µl of NT2 buffer to remove unbound protein. The immobilized THAP-His6 was incubated with the random pool of double stranded 80 bp DNA (2 to 5µg) in 100 µl of Binding buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.05% NP-40, 0.5 mM EDTA, 100 µg/ml BSA, and 20 to 50 µg of poly(dI-dC)) for 10 minutes at room temperature. The beads were then washed 6 times with 500 µl of NT2 buffer. The protein/DNA complex were then subjected to extraction with phenol/chloroform and precipitation with ethanol using 10 µg of

glycogen as a carrier. About one fifth of the recovered DNA was then amplified by 15 to 20 cycles of PCR and used for the next round of selection. After 8 rounds of selection, the NaCl concentration was progressively increased to 150 mM.

After 12 rounds of selection by THAP-His6, pools of amplified oligonucleotides were digested with Xba I and Hind III and cloned into pBluescript II KS - (Stratagene) and individual clones were sequenced using Big Dye terminator Kit (Applied Biosystem).

The results of the sequence analysis show that the THAP domain of human THAP1 is a site-specific DNA binding domain. Two consensus sequences were deduced from the alignment of two sets of nucleotide sequences obtained from the above SELEX procedure (each set containing 9 nucleic acid sequences). In particular, it was found that the THAP domain recognizes GGGCAA or TGGCAA DNA target sequences preferentially organized as direct repeats with 5 nucleotide spacing (DR-5 motifs). The consensus sequence being GGGCAAnnnnnTGGCAA (SEQ ID NO: 149). Additionally, THAP recognizes everted repeats with 11 nucleotide spacing (ER-11 motifs) having a consensus sequence of TTGCCAnnnnnnnnnnnGGGCAA (SEQ ID NO: 159). Although GGGCAA and TGGCAA sequences constitute the preferential THAP domain DNA binding sites, GGGCAT, GGGCAG and TGGCAG sequences are also DNA target sequences recognized by the THAP domain.

EXAMPLE 29

High throughput in vitro assay to identify inhibitors of THAP-family polypeptide or THAP-family interactions with nonspecific DNA targets

High throughput assays for the detection and quantification of THAP1-nonspecific DNA binding is carried out using a scintillation proximity assay. Materials are available from Amersham (Piscataway, NJ) and assays can be carried out according to Gal S. et al, 6th Ann. Conf. Soc. Biomol. Screening, 6-9 Sept 2000, Vancouver, B.C.), the disclosure of which is incorporated herein by reference in its entirety.

Random double stranded DNA probes are prepared and labelled using [^{3}H]TTP and terminal transferase to a suitable specific activity (e.g. approx. 420i/mmol). THAP1 protein or a portion thereof is prepared and the quantity of THAP1 protein or a portion thereof is determined via ELISA. For assay development purposes,

electrophoretic mobility shift assays (EMSA) can be carried out to select suitable assay parameters. For the high throughput assay, ^{3}H labelled DNA, anti-THAP1 monoclonal antibody and THAP1 in binding buffer (Hepes, pH 7.5; EDTA; DTT; 10 mM ammonium sulfate; KCl and Tween-20) are combined. The assay is configured in a standard 96-well plate and incubated at room temperature for 5 to 30 minutes, followed by the addition of 0.5 to 2 mg of PVT protein A SPA beads in 50-100 µl binding buffer. The radioactivity bound to the SPA beads is measured using a TopCount™ Microplate Counter (Packard Biosciences, Meriden, CT).

EXAMPLE 30

High throughput in vitro assay to identify inhibitors of THAP-family polypeptide or THAP-family interactions with specific DNA targets

High throughput assays for the detection and quantification of THAP1 specific DNA binding is carried out using a scintillation proximity assay. Materials are available from Amersham (Piscataway, NJ) and assays can be carried out according to Gal S. et al, 6th Ann. Conf. Soc. Biomol. Screening, 6-9 Sept 2000, Vancouver, B.C.).

THAP1-specific double stranded DNA probes corresponding to THAP1 DNA binding sequences obtained according to Example 20 are prepared. The probes are labelled using [^{3}H]TTP and terminal transferase to a suitable specific activity (e.g. approx. 420i/mmol). THAP1 protein or a portion thereof is prepared and the quantity of THAP1 protein or a portion thereof is determined via ELISA. For assay development purposes, electrophoretic mobility shift assays (EMSA) can be carried out to select suitable assay parameters. For the high throughput assay, ^{3}H labelled DNA, anti-THAP1 monoclonal antibody, 1µg non-specific DNA (double or single stranded poly-dAdT) and THAP1 protein or a portion thereof in binding buffer (Hepes, pH7.5; EDTA; DTT; 10 mM ammonium sulfate; KCl and Tween-20) are combined. The assay is configured in a standard 96-well plate and incubated at room temperature for 5 to 30 minutes, followed by the addition of 0.5 to 2 mg of PVT protein A SPA beads in 50-100µl binding buffer. The radioactivity bound to the SPA beads is measured using a TopCount™ Microplate Counter (Packard Biosciences, Meriden, CT).

EXAMPLE 31

Preparation of antibody compositions

Substantially pure THAP1 protein or a portion thereof is obtained. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms per ml. Monoclonal or polyclonal antibodies to the protein can then be prepared as follows: Monoclonal Antibody Production by Hybridoma Fusion Monoclonal antibody to epitopes in the THAP1 protein or a portion thereof can be prepared from murine hybridomas according to the classical method of Kohler and Milstein (Nature, 256: 495, 1975) or derivative methods thereof (see Harlow and Lane, Antibodies A Laboratory Manual, Cold Spring Harbor Laboratory, pp. 53-242, 1988), the disclosure of which is incorporated herein by reference in its entirety.

Briefly, a mouse is repetitively inoculated with a few micrograms of the THAP1 protein or a portion thereof over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., Meth. Enzymol. 70: 419 (1980), the disclosure of which is incorporated herein by reference in its entirety. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. et al. Basic Methods in Molecular Biology, Elsevier, New York., Section 21-2, the disclosure of which is incorporated herein by reference in its entirety.

Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogeneous epitopes in the THAP1 protein or a portion thereof can be prepared by immunizing suitable non-human animal with the THAP1 protein or a portion thereof, which can be unmodified or

modified to enhance immunogenicity. A suitable nonhuman animal, preferably a non-human mammal, is selected. For example, the animal may be a mouse, rat, rabbit, goat, or horse. Alternatively, a crude protein preparation which, has been enriched for THAP1 or a portion thereof can be used to generate antibodies. Such proteins, fragments or preparations are introduced into the non-human mammal in the presence of an appropriate adjuvant (e. g. aluminum hydroxide, RIBI, etc.) which is known in the art. In addition the protein, fragment or preparation can be pretreated with an agent which will increase antigenicity, such agents are known in the art and include, for example, methylated bovine serum albumin (mBSA), bovine serum albumin (BSA), Hepatitis B surface antigen, and keyhole limpet hemocyanin (KLH). Serum from the immunized animal is collected, treated and tested according to known procedures. If the serum contains polyclonal antibodies to undesired epitopes, the polyclonal antibodies can be purified by immunoaffinity chromatography.

Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987), the disclosure of which is incorporated herein by reference in its entirety. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al. J. Clin. Endocrinol. Metab. 33: 988-991 (1971), the disclosure of which is incorporated herein by reference in its entirety. Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. et al., Chap. 19 in: Handbook of Experimental Immunology D. Wier (ed) Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12: M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: Manual of Clinical Immunology, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D. C. (1980).

Antibody preparations prepared according to either the monoclonal or the polyclonal protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; or they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

EXAMPLE 32

Two Hybrid THAP1/Chemokine Interaction Assay

Two-hybrid interaction between THAP1 and chemokines CCL21, CCL19, CXCL9 and CXCL10 or cytokine IFNγ was tested by cotransformation of AH109 with pGADT7-THAP1 and pGBKT7-CCL21, -CCL19, -CXCL9, -CXCL10 and -IFNγ plasmids and selection of transformants by His and Ade double auxotrophy according to manufacturer's instructions (MATCHMAKER two-hybrid system 3, Clontech). pGBKT7-chemokine vectors were generated using cDNAs encoding the mature forms of human chemokines CCL21 (see example 15) (SLC polypeptide SEQ ID NO: 271, SLC cDNA SEQ ID NO: 272); CCL19 (amino acids 22-98 of GenBank Accession No. NM_006274) (CCL19 polypeptide SEQ ID NO: 273, CCL19 cDNA SEQ ID NO: 274); CXCL9 (amino acids 23-125 of GenBank Accession No. NM_002416) (CXCL9 polypeptide SEQ ID NO: 275, CXCL9 cDNA SEQ ID NO: 276) CXCL10 (amino acids 22-98 of GenBank Accession No. NM_001565) (CXCL10 polypeptide SEQ ID NO: 277, CXCL10 cDNA SEQ ID NO: 278) or cytokine IFNγ (amino acids 21-166 of GenBank Accession No. NM_000619) (IFNγ polypeptide SEQ ID NO: 279, IFNγ cDNA SEQ ID NO: 280), amplified by PCR, respectively, from Image clones No. 1707527 (hCCL19) with primers CCL19-1 (5'-GCGGAATCATGGGCACCAATGATGCTGAAGACTG-3') (SEQ ID NO: 281) and CCL19-2 (5'-GCGGGATCCTTAACTGCTGCGGCGCTTCATCTTG-3') (SEQ ID NO: 282), No. 5228247 (hCXCL9) with primers CXCL9-1 (5'-GCCGAATTCACCCCAGTAGTGAGAAAGGGTCGCTG-3') (SEQ ID NO: 283) and CXCL9-2 (5'-CGCGGATCCTTATGTAGTCTTCTTTTGACGAGAACGTTG-3') (SEQ ID NO: 284), No. 4274617 (hCXCL10) with primers CXCL10-1 (5'-GCCGAATTCGTACCTCTCTCTAGAACCGTACGCTGT-3') (SEQ ID NO. 285) and

CXCL10-2 (5'-GCGGGATCCTTAAGGAGATCTTTTAGACATTTCCTTGCTA-3') (SEQ ID NO. 286), No. 2403734 (hIFNγ) with primers IFN-1 (5'-GCGGAATCATGTGTTACTGCCAGGACCCATATG-3') (SEQ ID NO: 287) and IFN-2 (5'-GCGGGATCCTTACTGGGATGCTCTTCGACCTTG-3') (SEQ ID NO: 288). The PCR products were digested with EcoRI and BamHI, and cloned between EcoRI and BamHI cloning sites of vector pGBKT7 (Clontech). Positive two-hybrid interaction of THAP1 was observed with chemokines CCL21, CCL19, and CXCL9 while chemokine CXCL10 gave an intermediate result (+/-) in this two-hybrid assay (see Figure 19). The negative cytokine control, IFNγ, did not have a positive interaction.

It will be appreciated that the above-described methods can be used to determine whether any particular chemokine binds to any THAP-family polypeptide. For example, cDNAs encoding THAP-family members THAP1 to THAP11 as well as THAP0 from humans and other species can be cloned into a first component vector of a two hybrid system. cDNAs encoding chemokines can be cloned into a second component vector of a two hybrid system. The two vectors can be transformed into an appropriate yeast strain, wherein the polypeptides are expressed and tested for interaction. For example, chemokine CCL5 (polypeptide SEQ ID NO: 289, cDNA SEQ ID NO: 290) can be tested for interaction with full-length THAP-1 or particular portions of THAP1, such as a nested deletion series. Chemokines which can be tested for interaction with THAP-family proteins include, but are not limited to, XCL1, XCL2, CCL1, CCL2, CCL3, CCL3L1, SCYA3L2, CCL4, CCL4L, CCL5, CCL6, CCL7, CCL8, SCYA9, SCYA10, CCL11, SCYA12, CCL13, CCL14, CCL15, CCL16, CCL17, CCL18, CCL19, CCL20, CCL21, CCL22, CCL23, CCL24, CCL25, CCL26, CCL27, CCL28, clone 391, CARP CC-1, CCL1, CK-1, regakine-1, K203, CXCL1, CXCL1P, CXCL2, CXCL3, PF4, PF4V1, CXCL5, CXCL6, PPBP, SPBPBP, IL8, CXCL9, CXCL10, CXCL11, CXCL12, CXCL14, CXCL15, CXCL16, NAP-4, LFCA-1, Scyba, JSC, VHSV-induced protein, CX3CL1 and fCL1.

EXAMPLE 33

In Vitro THAP1/Chemokine Interaction Assay

To confirm the interaction observed in yeast two-hybrid system, we performed *in vitro* GST pull down assays. THAP1, expressed as a GST-tagged fusion protein and immobilized on glutathione sepharose, was incubated with radiolabeled chemokines that were translated *in vitro*.

To generate the GST-THAP1 expression vector, the full-length coding region of THAP1 (a nucleic acid encoding amino acids 1-213 of THAP1) was amplified by PCR from HEVEC cDNA with primers *2HMR8* (5'-CGCGGATCCGTGCAGTCCTGCTCCGCCTACGGC-3') (SEQ ID NO: 291 and *2HMR11* (5'-CCGAATTCTTATGCTGGTACTTCAACTATTTCAAAGTAG-3') (SEQ ID NO: 292), digested with BamHI and EcoRI, and cloned in frame downstream of the Glutathione S-Transferase ORF, between the BamHI and EcoRI sites of the pGEX-2T prokaryotic expression vector (Amersham Pharmacia Biotech, Saclay, France). The GST-THAP1 fusion protein encoded by plasmid pGEX-2T-THAP1 and the control GST protein encoded by plasmid pGEX-2T, were then expressed in *E.Coli* DH5α and purified by affinity chromatography with glutathione sepharose according to supplier's instructions (Amersham Pharmacia Biotech). The yield of proteins used in GST pull-down assays was determined by SDS-Polyacrylamide Gel Electrophoresis (PAGE) and Coomassie blue staining analysis.

In vitro-translated chemokines were generated with the TNT-coupled reticulocyte lysate system (Promega, Madison, WI, USA) using as templates pGBKT7-CCL21, -CCL19, -CXCL9 and -CXCL10 chemokine vectors (see Example 32) or pCMV-SPORT6-CCL5 plasmid (Image clone No. 4185200). *In vitro*-translated IFNγ cytokine was generated similarly using as template plasmid pGBKT7-IFNγ. A 25 µl volume of ^{35}S-labelled chemokine was incubated with immobilized GST-THAP1 or GST proteins overnight at 4 °C, in the following binding buffer: 10 mM NaPO4 pH 8.0, 140 mM NaCl, 3 mM MgCl2, 1 mM dithiothreitol (DTT), 0.05% NP40, and 0.2 mM phenylmethyl sulphonyl fluoride (PMSF), 1 mM Na vanadate, 50 mM β-glycerophosphate, 25 µg/ml chymotrypsine, 5 µg/ml aprotinin, and 10 µg/ml leupeptin. Beads were then washed 5 times in 1 ml binding buffer. Bound proteins were eluted with 2X Laemmli SDS-PAGE sample buffer, fractionated by 10% SDS–PAGE and

visualized by fluorography using Amplify (Amersham Pharmacia Biotech). GST/THAP1 specifically bound to chemokines CCL21, CCL19, CCL5, CXCL9 and CXCL10 but not cytokine IFNγ (Figures 19 and 20). Figure 19 shows that CCL21, CCL19, CCL5 and CXCL9 all strongly bound to immobilized GST-THAP1 (indicated by +++ in the column labelled "*In vitro* binding to GST-THAP1"). CXCL10 also bound to immobilized GST-THAP1 (indicated by ++ in the column labelled "*In vitro* binding to GST-THAP1"). The cytokine IFNγ did not bind to immobilized GST-THAP1 (indicated by - in the column labelled "*In vitro* binding to GST-THAP1"). Chemokines CCL21, CCL19, CCL5, CXCL9 and CXCL10 failed to interact with GST beads (negative control). Figure 20a shows that equivalent amounts of chemokine or cytokine were tested in the *in vitro* GST-THAP1 binding assays. Figure 20b shows that neither the cytokine, IFNγ, nor any of the chemokines bound to immobilized GST alone. Figure 20c shows that chemokines, CXCL10, CXCL9 and CCL19, but not the cytokine IFNγ, bound to immobilized GST-THAP1 fusions.

It will be appreciated that the above-described methods can be used to determine whether any particular chemokine binds to any THAP-family polypeptide. For example, cDNAs encoding THAP-family members THAP1 to THAP11 as well as THAP0 from humans and other species can be cloned and expressed as a GST fusion protein and immobilized to a solid support. cDNAs encoding chemokines can be translated *in vitro* and the resulting proteins incubated with the immobilized GST-THAP family fusions. Furthermore, a nested deletion series and/or point mutants of the THAP-family polypeptides can also be generated as GST-fusions and tested to determine the exact location of the chemokine binding domain for any THAP-family polypeptide with respect to any chemokine. Chemokines which can be tested for binding to THAP-family proteins include, but are not limited to, XCL1, XCL2, CCL1, CCL2, CCL3, CCL3L1, SCYA3L2, CCL4, CCL4L, CCL5, CCL6, CCL7, CCL8, SCYA9, SCYA10, CCL11, SCYA12, CCL13, CCL14, CCL15, CCL16, CCL17, CCL18, CCL19, CCL20, CCL21, CCL22, CCL23, CCL24, CCL25, CCL26, CCL27, CCL28, clone 391, CARP CC-1, CCL1, CK-1, regakine-1, K203, CXCL1, CXCL1P, CXCL2, CXCL3, PF4, PF4V1, CXCL5, CXCL6, PPBP, SPBPBP, IL8, CXCL9, CXCL10, CXCL11, CXCL12, CXCL14, CXCL15, CXCL16, NAP-4, LFCA-1, Scyba, JSC, VHSV-induced protein, CX3CL1 and fCL1.

EXAMPLE 34

Chemotaxis Bioassay: Inhibition of CCL21/CCL19-Induced Chemotaxis by THAP1 Oligomeric Forms

To demonstrate inhibition of CCL21/CCL19-induced chemotaxis by THAP1 oligomers, fresh lymphocytes and a human cell line, each of which expresses the CCL21/CCL19 receptor CCR7, are assayed for a chemotactic response to chemokines in the presence or absence of oligomeric THAP1. Lymphocytes are purified from fresh heparinized human blood or mouse lymph nodes and grown in RPMI 1640 glutamax I (Invitrogen GIBCO). HuT78 cells are obtained from American Tissue Type Culture Collection (Accession Number TIB-161) and grown in Iscove's modified Dulbecco's medium with 4 mM L-Glutamine adjusted to contain 1.5g/l sodium bicarbonate (Invitrogen GIBCO). Recombinant CCL21 and CCL19 human chemokines are obtained from commercial suppliers (for example, R&D or Chemicon).

Chemokine CCL21 or CCL19 is diluted in the appropriate culture medium without serum at 20 ng/ml and 75 µl of this solution is transferred in appropriated wells of a 96-well microplate. Recombinant human THAP1 oligomers (GST-THAP1 or Fc-THAP1 chimera) are serially diluted starting at 500 nM and 25 µl of the different dilutions are transferred in appropriate wells. Transwells are set carefully on each well and 100 µl of a cell suspension at 8.10^6 cell/ml is added in the upper chamber. Following a 4-hour incubation at 37°C and 5% CO_2, the cells which have migrated to the lower chamber are quantified using the Celltiter Glo system (Promega). A staining of the filter is also performed to verify that no cells adhered to the lower side of the filter after the migration. The degree of CCL21/CCL19 induced chemotaxis inhibition by THAP1 oligomeric forms is calculated by comparing the number of cells which have migrated in the presence or absence of the THAP1 oligomeric forms.

EXAMPLE 35

Inhibition of CCL21/CCL19-Induced Lymphocyte Adhesion to Endothelial Cells *In Vivo* by THAP1 Oligomeric Forms

The capacity of THAP1 oligomeric forms to block the activity of CCL21/CCL19 *in vivo*, including CCL21/CCL19-induced lymphocyte adhesion to endothelial cells, is assessed by measuring the 'rolling/sticking phenotype' of

lymphocytes in mouse lymph nodes HEVs (High endothelial venules) using intravital microscopy (microscopy on live animals) as described in von Andrian (1996) Microcirculation (3):287-300; and von Andrian UH, M'Rini C. (1998) Cell Adhes Commun. 6(2-3):85-96), the disclosures of which are incorporated herein by reference in their entireties. The rolling/sticking assay is performed as follows. In brief, the different steps of lymphocyte migration through HEVs (tethering, rolling, sticking, transendothelial migration) are analyzed by intravital microscopy in mice treated with recombinant human THAP1 oligomers (GST-THAP1 or Fc-THAP1 chimera). For observation of lymph nodes, HEVs vessels (and adhesion processes occurring in these vessels) by intravital microscopy, a microsurgical exposition of the subiliac (superficial inguinal) lymph node is made on an anaesthetized mouse. Briefly, BALB/c mice (Charles River, St Germain sur l'Arbresle, France) are anesthetized by intraperitoneal injection of 5 mg/mL ketamine and 1 mg/mL xylasine (10 mL/kg) and surgically prepared under a stereomicroscope (Leica Microsystems SA, Rueil-Malmaison, France) to allow exposure of the node vessels. A catheter is inserted in the contralateral femoral artery to permit subsequent retrograde injections of fluorescent cell suspensions or recombinant THAP1 oligomeric forms (GST-THAP1 or Fc-THAP1, 10-100 µg in 250 µl saline injected and allowed to bind for 5–30 min before injection of fluorescent cell suspensions). The mouse is then transferred to an intravital microscope (INM 100; Leica Microsystems SA). Body temperature is maintained at 37°C using a padding heater. Lymph node vessels and fluorescent cells are observed through 10 × or 20 × water immersion objective (Leica Microsystems SA) by transillumination or epifluorescence illumination. Transilluminated and fluorescent events are visualized using a silicon-intensified target camera (Hamamatsu Photonics, Massy, France) and recorded for later off-line analysis (DSR-11 Sony, IEC-ASV, Toulouse). Lymphocyte behavior in lymph node vessels is analyzed off-line as previously described (von Andrian (1996) Microcirculation (3):287-300; and von Andrian UH, M'Rini C. (1998) Cell Adhes Commun. 6(2-3):85-96). Briefly, the rolling fraction is determined in every visible lymph node HEV as the percentage of lymphocytes interacting with the endothelial lining over the total cell number entering the venule during an observation period. Rolling cells that become subsequently adherent are included in the rolling fraction. The sticking fraction is determined as percentage of rollers that becomes firmly adherent in

HEVs for more than 20 seconds. Only vessels with more than 10 rolling cells are included. The degree of inhibition of CCL21/CCL19-induced lymphocyte adhesion by THAP1 oligomers *in vivo* is calculated by comparing the number of lymphocytes sticking to endothelial cells (sticking fractions) in the presence or absence of the THAP1 oligomeric forms.

EXAMPLE 36

Use of THAP1 Oligomeric Forms to Antagonize Chemokines in a Mouse Model of Rheumatoid Arthritis

This experiment is designed to test effect of antagonizing chemokines with THAP1 oligomeric forms in a mouse model of rheumatoid arthritis, the well-known collagen-induced arthritis model. In each experiment, male DBA/1 mice are immunized with collagen on day 21 and are boosted on day 0. Mice are treated daily from days 0-14 with IP injections of THAP1 oligomeric forms (GST-THAP1 or THAP1-Fc chimera) at 150, 50, 15, and 5 μg/day, and compared to mice treated with control proteins (GST or human IgG1), at 150 μg/day (n=15/group in each experiment). The incidence and severity of arthritis is monitored in a blind study. Each paw is assigned a score from 0 to 4 as follows: 0=normal; 1=swelling in 1 to 3 digits; 2=mild swelling in ankles, forepaws, or more than 3 digits; 3=moderate swelling in multiple joints; 4=severe swelling with loss of function. Each paw is totaled for a cumulative score/mouse. The cumulative scores are then totaled for mice in each group for a mean clinical score. Groups of 15 mice are treated with the indicated doses of THAP1-Fc or with 150 μg/day of human IgG1. The capacity of THAP1 oligomeric forms (GST-THAP1 or THAP1-Fc chimera) to reduce the disease incidence and severity of arthritis is determined by comparison with the control group.

EXAMPLE 37

Use of THAP1 Oligomeric Forms to Antagonize Chemokines in a Mouse of Inflammatory Bowel Disease

The effect of blocking chemokines with THAP1-Fc chimera is analyzed in an experimentally induced model of Inflammatory Bowel Disease (IBD). One of the most widely used models of IBD is the DSS model (dextran sulphate salt). In this model,

dextran sulphate salt (M.W. typically about 40,000 but molecular weights from 40,000 to 500,000 can be used) is given to mice (or other small mammals) in their drinking water at 2% to 8%. In some studies, C57BL/6 mice are given 2% DSS from day 0 to day 7 (D0 - D7), wherein the number of mice per group equals four (n=4). The study groups are divided as follows: No DSS + human IgG1 (250 μg/day/mouse D0 - D7); 2% DSS + THAP1-Fc (100 μg/day/mouse D0 - D7); 2% DSS + THAP1-Fc (250 μg/day/mouse D0 - D7); 2% DSS + THAP1-Fc (500 μg/day/mouse D0 - D7); 2% DSS + human IgG1 (250 μg/day/mouse D0 - D7). Mice are weighed daily. Weight loss is a clinical sign of the disease. Tissues are harvested at day 8 (D8). Histopathology is performed on the following tissues: small intestine, large intestine and mesenteric lymph nodes (MLN). The capacity of the THAP1-Fc chimera, to attenuate some of the weight loss associated with DSS-induced colitis, and to reduce inflammation in the large intestine is determined by comparing mice treated with THAP1-Fc to mice treated with control human IgG1.

The methods, compositions, and devices described herein are presently representative of preferred embodiments and are exemplary and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention and are defined by the scope of the disclosure. Accordingly, it will be apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention.

As used in the claims below and throughout this disclosure, by the phrase "consisting essentially of" is meant including any elements listed after the phrase, and limited to other elements that do not interfere with or contribute to the activity or action specified in the disclosure for the listed elements. Thus, the phrase "consisting essentially of" indicates that the listed elements are required or mandatory, but that other elements are optional and may or may not be present depending upon whether or not they affect the activity or action of the listed elements.

REFERENCES

Numerous literature and patent references have been cited in the present application. All references cited are incorporated by reference herein in their entireties.

For some references, the complete citation is in the body of the text. For other references the citation in the body of the text is by author and year, the complete citation being as follows:

Alcalay, M., Tomassoni, L., Colombo, E., Stoldt, S., Grignani, F., Fagioli, M., Szekely, L., Helin, K., and Pelicci, P. G. (1998). The promyelocytic leukemia gene product (PML) forms stable complexes with the retinoblastoma protein. Mol Cell Biol *18*, 1084-93.

Arcone, et al., Nucl. Acids Res., 16(8): 3195-3207, 1988.

Baichwal and Sugden, In: Gene Transfer, Kucherlapati R, ed., New York, Plenum Press, pp. 117-148, 1986.

Barradas, M., Monjas, A., Diaz-Meco, M. T., Serrano, M., and Moscat, J. (1999). The downregulation of the pro-apoptotic protein Par4 is critical for Ras-induced survival and tumor progression. Embo J *18*, 6362-9.

Bartlett et al., Proc. Nat Acad Sci. USA, 93:8852-8857, 1996.

Benvenisty and Neshif, Proc. Nat Acad Sci. USA, 83:9551-9555, 1986.

Berra, E., Municio, M. M., Sanz, L., Frutos, S., Diaz-Meco, M. T., and Moscat, J. (1997). Positioning atypical protein kinase C isoforms in the UV-induced apoptotic signaling cascade. Mol Cell Biol *17*, 4346-54.

Bloch, D. B., Chiche, J. D., Orth, D., de la Monte, S. M., Rosenzweig, A., and Bloch, K. D. (1999). Structural and functional heterogeneity of nuclear bodies. Mol Cell Biol *19*, 4423-30.

Boghaert E.R. et al, Cell Growth Differ. 1997 Aug; 8(8):881-90

Burge, C., Karlin, S. (1997) Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. 268, 78-94

Carter and Flotte, Ann. N. Y Acad Sci., 770:79-90, 1995.

Chattedee, et al., Ann. N. Y Acad Sci., 770:79-90, 1995.

Chen and Okayama, Mol. Cell Biol., 7:2745-2752, 1987.

Coffin, In: Virology, ed., New York: Raven Press, pp. 143 7-15 00, 1990.

Coupar et al., Gene, 68:1-10, 1988.

Dani, et al., J Biol. Chem., 264:10119-10125, 1989.

Diaz-Meco, M. T., Municio, M. M., Frutos, S., Sanchez, P., Lozano, J., Sanz, L., and Moscat, J. (1996). The product of Par4, a gene induced during apoptosis, interacts selectively with the atypical isoforms of protein kinase C. Cell *86*, 777-86.

Dubensky et al., Proc. Nat'l Acad Sci. USA, 81:7529-7533, 1984.

Fechheimer et al., Proc. Nat Acad. Sci. USA, 84:8463-8467, 1987.

Felsenstein, J. 1989. PHYLIP -- Phylogeny Inference Package (Version 3.2). *Cladistics* 5: 164-166.

Ferkol et al., FASEB J., 7:1081-1091, 1993.

Ferrari et al., J Virol., 70:3227-3234, 1996.

Fisher et al., J Virol., 70:520-532, 1996.

Flotte et al., Proc. Nat'l Acad. Sci. USA, 90:10613-10617, 1993.

Fogal, V., Gostissa, M., Sandy, P., Zacchi, P., Sternsdorf, T., Jensen, K., Pandolfi, P. P., Will, H., Schneider, C., and Del Sal, G. (2000). Regulation of p53 activity in nuclear bodies by a specific PML isoform. Embo J *19*, 6185-95.

Fraley et al., Proc. Nat'l Acad. Sci. USA, 76:3348-3352, 1979.

Friedmann, "Progress toward human gene therapy," Science, 244:1275-1281, 1989.

Ghosh and Bachhawat, In: Liver diseases, targeted diagnosis and therapy using specific receptors and ligands, (Wu G, Wu C ed,), New York: Marcel Dekker, pp. 87 104, 1991.

Girard, J. P., Baekkevold, E. S., Feliu, J., P, B., and Amalric, F. (1999). Molecular cloning and functional analysis of SUT-1, a sulfate transporter from human high endothelial venules. Proc Natl Acad Sci USA *96:12772-12777.*

Girard, J. P., Baekkevold, E. S., Yamanaka, T., Haraldsen, G., Brandtzaeg, P., and Amalric, F. (1999). Heterogeneity of endothelial cells : the specialized phenotype of human high endothelial venules characterized by suppression subtractive hybridization. Am J Pathol *155:2043-2055.*

Girard, J. P., and Springer, T. A. (1995). Cloning from purified high endothelial venule cells of hevin, a close relative of the antiadhesive extracellular matrix protein SPARC. Immunity *2*, 113-123.

Goodman et al., Blood, 84:1492-1500, 1994.

Gopal, Mol. Cell Biol., 5:1188-1190, 1985.

Gossen and Bujard, Proc. Nat Acad. Sci. USA, 89:5547-5551, 1992.

Gossen et al., Science, 268:1766-1769, 1995.

Graham and Van Der Eb, Virology, 52:456-467, 1973.

Guo, Q., Fu, W., Xie, J., Luo, H., Sells, S. F., Geddes, J. W., Bondada, V., Rangnekar, V. M., and Mattson, M. P. (1998). Par4 is a mediator of neuronal degeneration associated with the pathogenesis of Alzheimer disease. Nat Med *4*, 957-62.

Hay et al., J Mol. Biol., 175:493 -510, 1984.

Hearing and Shenk, J Mol. Biol. 167:809-822, 1983.

Hearing et al., J Virol., 67:2555-2558, 1987.

Higgins, D. G., Thompson, J. D. and Gibson, T. J. (1996) Using CLUSTAL for multiple sequence alignments. Methods Enzymol., 266, 383-402.

Horowitz, Yandell, Park, Canning, Whyte, Buchkovich, Harlow, Weinberg, Dryja, "Point mutational inactivation of the retinoblastoma antioncogene," Science 243, 937 940,1989.

Hunt et al., Proc. Nat Acad. Sci. USA, 83:3786-3790, 1986.

Ishov, A. M., Sotnikov, A. G., Negorev, D., Vladimirova, O. V., Neff, N., Kamitani, T., Yeh, E. T., Strauss, J. F., 3rd, and Maul, G. G. (1999). PML is critical for ND10 formation and recruits the PML-interacting protein daxx to this nuclear structure when modified by SUMO-1. J Cell Biol *147*, 221-34.

Jareborg N, Birney E, Durbin R. Genome Res. 1999 Sep;9(9):815-24.

Jentsch, S., and Pyrowolakis, G. (2000). Ubiquitin and its kin: how close are the family ties? Trends Cell Biol *10*, 335-42.

Johnstone, R. W., See, R. H., Sells, S. F., Wang, J., Muthukkumar, S., Englert, C., Haber, D. A., Licht, J. D., Sugrue, S. P., Roberts, T., Rangnekar, V. M., and Shi, Y. (1996). A novel repressor, Par4, modulates transcription and growth suppression functions of the Wilms' tumor suppressor WT1. Mol Cell Biol *16*, 6945-56.

Joki et al., Human Gene Ther., 6:1507-1513, 1995.

Kageyama, et al., J Biol. Chem., 262(5):2345-2351, 1987.

Kaneda et al., Science, 243:375-378, 1989.

Kaplitt et al., Nat Genet., 8:148-153, 1994.

Kato et al, J Biol. Chem., 266:3361-3364, 1991.

Kessler et al., Proc. Nat Acad. Sci. USA, 93:14082-14087, 1996.

Klein et al., Nature, 327:70-73, 1987.

Koeberl et al., Proc. Nat 7 Acad. Sci. USA, 94:1426-1431, 1997.

Korhonen et al., Blood, 86:1828-1835, 1995.

Lallemand-Breitenbach, V., Zhu, J., Puvion, F., Koken, M., Honore, N., Doubeikovsky, A., Duprez, E., Pandolfi, P. P., Puvion, E., Freemont, P., and de The, H. (2001). Role of Promyelocytic Leukemia (PML) Sumolation in Nuclear Body Formation, 11S

Proteasome Recruitment, and As(2)O(3)-induced PML or PML/Retinoic Acid Receptor alpha Degradation. J Exp Med *193*, 1361-72.

LaMorte, V. J., Dyck, J. A., Ochs, R. L., and Evans, R. M. (1998). Localization of nascent RNA and CREB binding protein with the PML-containing nuclear body. Proc Natl Acad Sci U S A *95*, 4991-6.

Levrero et al., Gene, 10 1: 195-202, 199 1.

Li, H., Leo, C., Zhu, J., Wu, X., O'Neil, J., Park, E. J., and Chen, J. D. (2000). Sequestration and inhibition of Daxx-mediated transcriptional repression by PML. Mol Cell Biol *20*, 1784-96.

Mann et al., Cell, 33:153-159, 1983.

Mattson, M. P., Duan, W., Chan, S. L., and Camandola, S. (1999). Par4: an emerging pivotal player in neuronal apoptosis and neurodegenerative disorders. J Mol Neurosci *13*, 17-30.

Maul, G. G., Negorev, D., Bell, P., and Ishov, A. M. (2000). Review: properties and assembly mechanisms of ND10, PML bodies, or PODs. J Struct Biol *129*, 278-87.

Melchior, F. (2000). SUMO--nonclassical ubiquitin. Annu Rev Cell Dev Biol *16*, 591-626.

Mizukami et al., Virology, 217:124-130, 1996.

Muller, S., Matunis, M. J., and Dejean, A. (1998). Conjugation with the ubiquitin-related modifier SUMO-1 regulates the partitioning of PML within the nucleus. Embo J *17*, 61-70.

McCown et al., Brain Res., 713:99-107, 1996.

Nicolas and Rubenstein, In: Vectors: A survey of molecular cloning vectors and their uses, Rodriguez and Denhardt (eds.), Stoneham: Butterworth, pp. 493-513, 1988.

Nicolau and Sene, Biochim. Biophys. Acta, 721:185-190, 1982.

Nicolau et al., Methods Enzymol., 149:157-176, 1987.

Olivierio, et al., EMBO J, 6(7):1905-1912, 1987.

Page, G., Kogel, D., Rangnekar, V., and Scheidtmann, K. H. (1999). Interaction partners of Dlk/ZIP kinase: co-expression of Dlk/ZIP kinase and Par4 results in cytoplasmic retention and apoptosis. Oncogene *18*, 7265-73.

Pape and Kim, Mol. Cell. Biol., 974-982, 1989.

Paskind et al., Virology, 67:242-248, 1975.

Perales et al., Proc. Nat Acad Sci. 91:4086-4090, 1994.

Ping et al., Microcirculation, 3:225-228, 1996.

Poli and Cortese, Proc. Nat'l Acad Sci. USA, 86:8202-8206, 1989.

Potter et al., Proc. Nat Acad. Sci. USA, 81:7161-7165, 1984.

Prowse and Baumann, Mol Cell Biol, 8(l):42-51, 1988.

Quignon, F., De Bels, F., Koken, M., Feunteun, J., Ameisen, J. C., and de The, H. (1998). PML induces a novel caspase-independent death process. Nat Genet *20*, 259-65.

Radler et al., Science, 275:810-814, 1997. Rb Protein Inhibits p84N5-induced Apoptosis, 10:3261, October 1999.

Ridgeway, In: Vectors: A survey ofmolecular cloning vectors and their uses, Rodriguez RL, Denhardt DT, ed., Stoneham:Butterworth, pp. 467-492, 1988.

Rippe et al., Mol. Cell Biol., 10:689-695, 1990.

Ron et al., Mol. Cell. Biol., 2887-2895, 199 1.

Roux et al., Proc. Nat Acad. Sci. USA, 86:9079-9083, 1989.

Ruggero, D., Wang, Z. G., and Pandolfi, P. P. (2000). The puzzling multiple lives of PML and its role in the genesis of cancer. Bioessays *22*, 827-35.

Sells, S. F., Han, S. S., Muthukkumar, S., Maddiwar, N., Johnstone, R., Boghaert, E., Gillis, D., Liu, G., Nair, P., Monnig, S., Collini, P., Mattson, M. P., Sukhatme, V. P., Zimmer, S. G., Wood, D. P., Jr., McRoberts, J. W., Shi, Y., and Rangnekar, V. M. (1997). Expression and function of the leucine zipper protein Par4 in apoptosis. Mol Cell Biol *17*, 3823-32.

Speigelman, et al., J Biol. Chem., 264(3), 1811-1815, 1989.

Sternsdorf, T., Jensen, K., Reich, B., and Will, H. (1999). The nuclear dot protein sp100, characterization of domains necessary for dimerization, subcellular localization, and modification by small ubiquitin-like modifiers. J Biol Chem *274*, 12555-66.

Tartaglia, Ayres, Wong, and Goeddel, "A novel domain within the 55 kd TNF receptor signals cell death," Cell 74, 845-853, 1993.

Temin, In: Gene Transfer, Kucherlapati (ed.), New York: Plenum Press, pp. 149-188, 1986.

Tibbetts Cell, 12:243-249, 1977.

Tur-Kaspa et al., Mol. Cell Biol., 6:716-718, 1986.

Wagner et al., Proc. Nat'l Acad. Sci. 87(9):3410-3414, 1990.

Walther and Stein, J Mol. Med, 74:3 79-3 92, 1996.

Wang, Z. G., Delva, L., Gaboli, M., Rivi, R., Giorgio, M., Cordon-Cardo, C., Grosveld, F., and Pandolfi, P. P. (1998). Role of PML in cell growth and the retinoic acid pathway. Science *279*, 1547-51.

Wang, Z. G., Ruggero, D., Ronchetti, S., Zhong, S., Gaboli, M., Rivi, R., and Pandolfi, P. P. (1998). PML is essential for multiple apoptotic pathways. Nat Genet *20*, 266-72.

Watt et al., Proc. Nat'l Acad Sci., 83(2): 3166-3170, 1986.

Wilson, et al., Mol. Cell. Biol., 6181-6191, 1990.

Wong et al., Gene, 10:87-94, 1980.

Wu and Wu, Adv. Drug Delivery Rev., 12:159-167, 1993.

Wu and Wu, Biochem., 27:887-892,1988.

Wu and Wu, J Biol. Chem., 262:4429-4432,1987.

Xiao et al., J Virol., 70:8098-8108, 1996.

Yang et al., Proc. Nat Acad. Sci USA, 87:9568-9572, 1990.

Yeh, E. T., Gong, L., and Kamitani, T. (2000). Ubiquitin-like proteins: new wines in new bottles. Gene *248*, 1-14.

Zechner, et al., Mol. Cell. Biol., 23 94-2401, 198 8.

Zhong, S., Hu, P., Ye, T. Z., Stan, R., Ellis, N. A., and Pandolfi, P. P. (1999). A role for PML and the nuclear body in genomic stability. Oncogene *18*, 7941-7.

Zhong, S., Muller, S., Ronchetti, S., Freemont, P. S., Dejean, A., and Pandolfi, P. P. (2000). Role of SUMO-1-modified PML in nuclear body formation. Blood *95*, 2748-52.

Zhong, S., Salomoni, P., and Pandolfi, P. P. (2000). The transcriptional role of PML and the nuclear body. Nat Cell Biol *2*, E85-90.

Zhong, S., Salomoni, P., Ronchetti, S., Guo, A., Ruggero, D., and Pandolfi, P. P. (2000). Promyelocytic leukemia protein (PML) and Daxx participate in a novel nuclear pathway for apoptosis. J Exp Med *191*, 631-40.